

M

Machine Failure Monitoring (MFM)

- ▶ [Journal Bearing Failure Monitoring \(JBFM\)](#)

Macro-cracks

- ▶ [Growth Characteristics of Large Fatigue Cracks](#)

Macropitting

- ▶ [Gear Surface Pitting Failure and Pitting Life Analysis](#)
- ▶ [Rolling Bearing Contact Fatigue](#)

Macropitting Resistance

- ▶ [Gear Surface Pitting Failure and Pitting Life Analysis](#)
- ▶ [Rolling Bearing Contact Fatigue](#)

Macroscopic Fatigue Cracks

- ▶ [Growth Characteristics of Large Fatigue Cracks](#)

Magnetic-Fluid Seals

- ▶ [Ferrofluid Seals](#)

Makeup

- ▶ [Lubricant Formulation](#)

Manufacturing Tribology

STEVEN R. SCHMID

Department of Aerospace and Mechanical Engineering,
University of Notre Dame, Notre Dame, IN, USA

Synonyms

[Process tribology](#)

Definition

Manufacturing tribology is the sub-discipline of tribology that addresses specific issues in manufacturing processes.

Scientific Fundamentals

Manufacturing as a science depends on the fusion of many specializations, one of the most important of which is tribology. However, while an understanding of tribology is essential for most manufacturing processes, and manufacturing is essential for modern society, the combination of manufacturing and tribology in an integrated textbook, or even organized course, is very rare. For the reader who requires additional information, they are directed to the classic manufacturing texts by Kalpakjian and Schmid (2008, 2010), and Schey (1999), and especially the remarkable text by Schey (1984), which is still the most comprehensive and most often-cited book on tribology in manufacturing despite its age. Schmid and Wilson (2001) have summarized tribological concerns in manufacturing, and generally expands on all of the issues raised here.

Tribology in manufacturing is often very different from tribology with machine elements. In cutting and

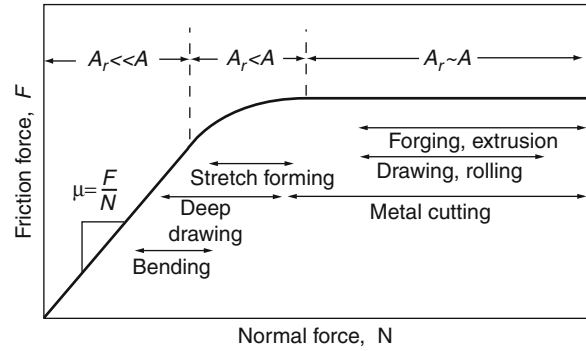
forming operations, the pressure in the tool-workpiece contact often exceeds the yield stress of the workpiece material. This renders the traditional theories of contact and lubrication of limited value. Special mechanisms of friction and wear may also be active and constraints on lubrication may apply. Thus, while the basic principles used in the tribology of machine elements are still valid, great caution must be used in applying them to the systems used in manufacturing.

This chapter is intended as an introduction to manufacturing tribology, and will introduce specific process considerations with emphasis on differences between machine element and manufacturing tribology.

General Tribology Concepts

Some of the most important fundamental tribology concepts that are difficult to apply to manufacturing are the following:

- Friction does not, in general, follow the three laws of friction associated with Coulomb friction. For example, in bulk deformation processes, friction force is not proportional to normal force; a coefficient of friction is not a constant; and friction is dependent on sliding speed for a number of cases and materials. This can be understood by recognizing that with machinery elements such as bearings or gears, asperity contact is rare and contact patches are isolated from one another even under unlubricated conditions. In manufacturing processes, contact patches are large, and the real contact area approaches the apparent contact area, as shown in Fig. 1.
- Surfaces are dynamic in manufacturing processes. Workpiece surfaces roughen due to unconstrained plastic deformation. The roughening can be attributed to Lüders bands or slip lines within a grain, but are more commonly caused by surface grain translation and rotation. Surface roughening due to plastic deformation is commonly referred to as “orange peel” because of the rough grainy appearance of metals and the similarity to the appearance of the surface of an orange. On the other hand, smooth tooling can flatten asperities to make a workpiece smoother. Also, machining and other material removal processes can remove a rough (or smooth) surface to expose nascent material and fundamentally change the surface chemistry. Thus, the surface roughness, texture, lay, and chemistry can all change within a process.
- The concepts of a film parameter and regimes of lubrication are certainly applicable to manufacturing



Manufacturing Tribology, Fig. 1 Friction force as a function of normal force and range of application for assorted manufacturing processes. The ranges shown are for unlubricated cases (Kalpakjian and Schmid 2008)

processes. The main difference is in typical values of film parameter and lubrication regime that are used. In most lubricated machine element applications, it is desired to achieve partial or full film lubrication. However, in manufacturing it is important to maintain some tooling intimacy with the workpiece in order to minimize orange peel. This requires that lubrication be restricted to mixed or boundary lubrication for most applications. Boundary lubricants in general have much less time to establish themselves on workpieces, and therefore are formulated with both workpiece and tooling in mind. Additive packages are in general more aggressive than in machine element applications, and include additional additives such as biocides, brighteners, and emulsifiers.

- With machine elements, wear is generally thought of as detrimental, but it is the important enabling mechanism for processes such as grinding, polishing, ultrasonic machining, and deburring operations. Wear is detrimental in cases where tooling, inserts, or dies are limited in life, to the detriment of process economics.
- Temperatures in manufacturing are much higher, in general, than in machine element applications. For example, isothermal forging of nickel-based superalloys can occur at 1,000°C, the peak temperature on a cutting tool insert can be 750°C, and hot rolling of steel routinely takes place at 400°C.

Bulk Deformation Processes

Bulk deformation includes forging, rolling, extrusion, drawing, swaging, and associated processes. These

processes involve plastic deformation throughout the workpiece volume. Some of the special concerns of bulk deformation processes are the following:

- Sheu and Wilson (1994) suggest the asperity hardness can be expressed by the relationship

$$H = \frac{2}{(f_1(\alpha)E + f_2(\alpha))} \quad (1)$$

where H is the effective hardness of the workpiece:

$$H = \frac{p_a - p_b}{k} \quad (2)$$

E is the dimensionless strain rate given by

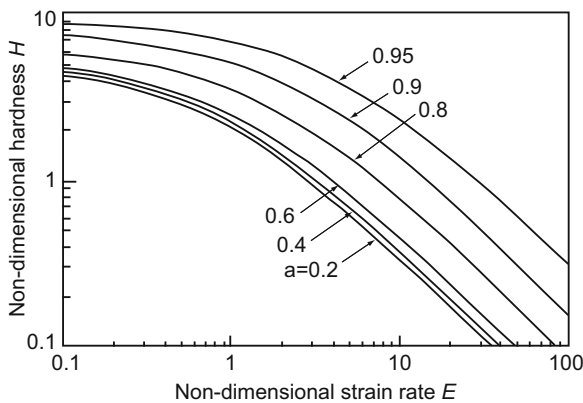
$$E = \frac{\dot{\epsilon}l}{v_f} \quad (3)$$

a is the indenter area ratio, $a = a/l$ and f_1 and f_2 are defined by

$$f_1(a) = 0.515 + 0.345a - 0.860a^2 \quad (4)$$

$$f_2 = \frac{1}{(2.571 - \alpha - \alpha \ln(1 - \alpha))} \quad (5)$$

The effective hardness given by Eq. 3 is shown in Fig. 2. An extremely important trend can be seen: the hardness of the workpiece asperities falls dramatically in the presence of substrate straining. This means that the rate at which asperities flatten will be much higher in metal working than in machine design applications. This has far-reaching implications and fundamentally changes the contact mechanics that occurs at asperity



Manufacturing Tribology, Fig. 2 Plot of asperity hardness as a function of substrate strain rate (Sheu and Wilson 1984)

scales. Because the effective hardness of plastically deforming workpieces is very low, the surface roughness of the tooling can be imparted to the workpiece (as is done, for example, with beverage cans or sheet metal parts such as automotive body panels). Also, the fractional contact area will be much larger than in machine element contacts, thereby changing the mechanics of friction and wear.

- The preferred lubricant for many hot and cold working operations, as well as cutting and grinding operations discussed below, is an emulsion, consisting of an insoluble oil phase suspended in water. Emulsions are much more widely used in manufacturing than in machine element applications, and display a complicated rheology.
- For hot working, lubricants include graphite (often as a graphite-in-oil or graphite-in-water suspension), soft metals or polymers, glass, or an emulsion. It should be noted that glass is an effective liquid lubricant at the elevated operating temperatures typical of many bulk forming processes, and its formulation can be modified to achieve a desired viscosity.
- The Hamrock-Dowson equations are well known as a design aide in that they predict the lubricant film thickness in elastohydrodynamic contacts. The equation of similar importance and utility in metal forming is the Wilson and Walowit (1972) equation, initially derived for rolling but variants of which apply to most metal forming operations. The Wilson-Walowit equation is

$$h = \frac{6\eta\gamma U}{\tan \theta (1 - e^{-\gamma\sigma})} \quad (6)$$

where h is the film thickness, η is the lubricant viscosity, α is the viscosity pressure coefficient, U is the mean inlet rolling speed, and θ is the contact angle between roll and workpiece. It should be noted that the product $\alpha\sigma$ is usually a large number; the denominator term in parentheses is usually very close to unity for most lubricants.

- Conversion coatings and roughening of surfaces is done to improve lubrication. With machinery elements, an improved lubrication condition generally results from smoothing or improved polishing of surfaces. Since manufacturing generally involves boundary or mixed lubrication, entrainment of lubricant by surfaces is often essential, and highly polished surfaces are costly and ineffective.
- Surface chemistry is less uniform than in machinery element applications. For example, an oxide on a steel surface may persist during a contact event on a rolling

element bearing, but the oxide on a steel workpiece will by necessity be broken up and spread apart during bulk deformation. The spaces between oxide patches will be filled by nascent material (from the substrate), which will be oxide- and boundary lubricant-free and will potentially have a very different microstructure than the surface layers. In general, the nascent material is much more aggressive from a wear standpoint and leads to higher friction.

- In rolling, some friction is needed (when no front tension is present) in order to pull the workpiece past the rolls and avoid a slick mill condition. However, high friction leads to excessive roll force (and may exacerbate roll flattening and bending) and torque, so that control, not elimination, of friction is the goal.
- It is not unusual in hot forging to have the workpiece preheated to 700°C and the tooling preheated to 300°C or so. Clearly, there is significant heat transfer from the workpiece to the tooling, and the workpiece surface layers are quickly cooled as a result. This increases the local stiffness and strength of the material, especially in surface layers or thin cross-sections. The effect of cooling is the same as if the friction coefficient were suddenly increased (higher tooling forces or power required, restricted workpiece deformation, etc.), but is not related to friction at all. The evolution of heat transfer and friction boundary conditions during a process is unavoidable and complicates simulation.
- In forging, it is common that the tooling approaches the workpiece with a normal velocity, and that lubricant films are generated by a normal squeeze effect. This is rare with machinery elements.
- In extrusion, it is common practice to place a glass or glass fiber pad at the front of the billet. As the hot billet is pushed through a die, it melts a layer of the glass that is then entrained as a lubricating film. However, at startup, the lubricant entrainment velocity is one half the velocity of the workpiece, and a start-stop-reverse (or incremental) motion is necessary.
- When simple sketches of extrusion and drawing operations are made, the only apparent difference is that the workpiece is pushed through dies in extrusion, while it is pulled through the dies in drawing. In actuality, the processes have very real differences, perhaps the most important being that extrusion is a batch process, while drawing can be a continuous process. More subtly, extrusion will use far greater reductions in area per pass than drawing, and therefore will generally be performed at a higher temperature to obtain greater workpiece ductility.

Good general references on extrusion are Saha (1999), Schey (1985), and Laue et al. (1981). General drawing references are Schey (1985) and Tassi (1981).

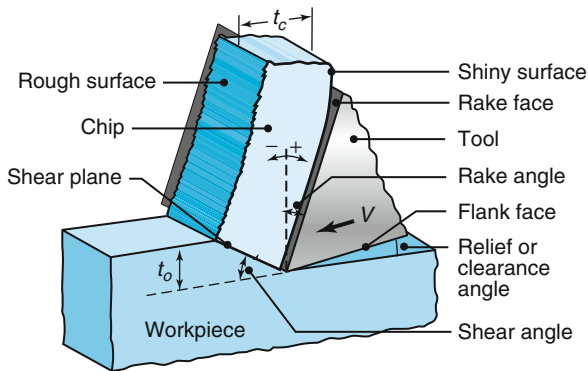
Sheet Metal Forming

Rolled products in the form of sheet metal are widely available, of high quality, and low cost. A well-developed manufacturing strategy is to form or stamp parts from sheet metal base stock. Sheet metal forming has similar concerns as bulk forming, in that orange peel is a serious concern, lubrication is more complicated than in machinery element situations, surfaces are dynamic, and boundary or mixed lubrication are the most common lubrication regimes. Some additional tribological concerns in sheet forming are:

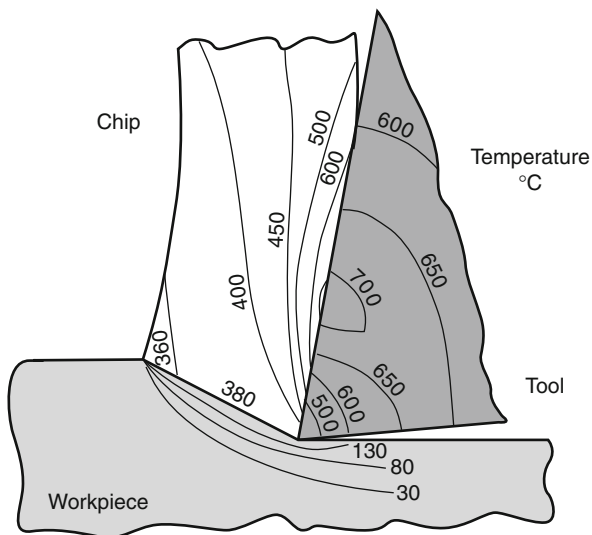
- With sheet metal forming, friction is classically not well described by either the Coulomb or Tresca friction laws. Instead, friction is a more complicated phenomenon.
- The final strain achieved by a workpiece, as well as its formability without fracture, are very dependent on the friction and hence the lubricant film thickness achieved.
- Surface finish in sheet metal products is often the most stringent. For example, mirror-like surfaces are required on aluminum beverage containers, which are produced in the hundreds of millions in the United States alone every day.
- Recently, new steel alloys have been developed for structural applications, especially in the automotive industry. Examples are TRIP (transformation-induced plasticity) and TWIP (twinning-induced plasticity) steels, which have very high strengths. To reduce forming forces, pressures, stiffness, and springback, these materials are processed through hot stamping. Thus, the thermal and friction concerns associated with hot bulk forming exist in hot stamping as well.

Cutting and Finishing

A simple orthogonal cutting operation is shown in Fig. 3. From the image, it can be seen that there is intense shear along a plane of the workpiece; this observation is the basis of the Ernst and Merchant shear plane theory and is the classic start of machining analysis. For the purposes of this chapter, it should be recognized as a heat source, as should the frictional interface between the chip and the tool. The main result is that the cutting tool temperatures can be very high, as shown in Fig. 4. Thus, although the wear mechanisms to be discussed in this chapter are also



Manufacturing Tribology, Fig. 3 Schematic illustration of orthogonal metal cutting (Kalpakjian and Schmid 2010). The rake face of the tool encounters elevated temperatures, severe abrasion, adhesion, thermal fatigue, and diffusion, and does not promote lubricant entrainment. It is perhaps the harshest tribological interface



Manufacturing Tribology, Fig. 4 Illustration of temperature distribution in metal cutting (Kalpakjian and Schmid 2010)

present in gears, bearings, and the like, the superposition of high temperatures makes them much more serious.

Metal cutting places extreme demands on tool materials. In addition to the very high pressures and temperatures in metal cutting, the geometries of the process often preclude the introduction of lubricant. Process geometries tend to become extremely complicated, and it is difficult to obtain closed form solutions for temperature

distributions, film thicknesses, etc. An excellent summary of the mechanical and tribological concerns in cutting is given by Shaw (1984).

Flank wear results in smearing of the workpiece surface by the cutting tool and, when excessive, results in poor workpiece surface finish. Based on flank wear considerations, the life of a cutting tool can be estimated from the Taylor tool life equation:

$$VT^n = \text{constant} \quad (7)$$

where V is the cutting speed, T is the tool life, and n is a constant of the cutting tool material. Since n is usually around 0.2 or so, speed is seen to have a drastic effect on tool life. The Taylor tool life is often generalized as

$$TV^{1/n}t^{1/m}b^{1/l} = \text{constant} \quad (8)$$

Where t is feed rate and b is depth of cut. Usually, $n < m < l$ (Shaw 1984).

Nose wear involves the loss of material from the cutting tool tip. Nose wear is detrimental because it results in higher cutting forces. Nose wear also has a tendency to promote unstable built-up edge, which consists of deposited workpiece material that periodically breaks free, resulting in compromise of workpiece surface finish.

According to the Archard wear law, the tool life can be maximized by using a tool material with high hardness. It is critical that the hardness be maintained at cutting temperatures, and also that it has a high fracture toughness in the case of interrupted cutting conditions or chatter. In addition, the tool should have chemical inertness at high temperatures and, in order to reduce operating temperatures, should have high thermal conductivity. Very few materials display all of these properties.

One of the common solutions to the material selection dilemma is to apply a coating of a hard and chemically inert material to a ductile and thermally conductive substrate. A relatively recent development is the production of multiphase coatings, where each layer is on the order of a few micrometers. Since hardness is known to increase with decreasing grain size, each layer has very high hardness.

Recent advances in cutting tool materials have allowed the continued increase in cutting speeds or life at a given cutting speed. Tool manufacturers have developed particular material formulations for the most commonly used alloys or series of alloys.

Finishing operations such as grinding, lapping, polishing, etc., use the abrasive action of small particles to remove material and refine surface finish. It is essential for the production of superior surfaces required for applications ranging from hard disk to bearing manufacture.

For the interested reader, the books by Shaw (1996) and Malkin (1989) are excellent resources.

Abrasives consist of small abrasive particles suspended in a polymer or ceramic matrix, referred to as resinoid or vitrified, respectively. When mounted onto a high-speed spindle and pushed against a surface, the particle can interact with the surface in a number of different ways:

- With high attack angle, the particle can generate a chip in the workpiece.
- At lower attack angle, a chip will not be generated, but the abrasive will plow a groove in the workpiece.
- When the contact forces are relatively low, the abrasive will merely interact elastically.

Compared with conventional machining processes, abrasive machining is much less efficient in terms of the energy required to remove a unit volume of material.

Key Applications

Tribology is important in most manufacturing operations. Friction is important in bulk forming operations since it affects strains achieved, and is essential to draw material through the rolls in metal rolling. Except for rolling and the flash region in impression forging, low friction is desired in order to reduce forming forces. Wear is important in tooling and dies in order to maximize life, but is also an essential phenomenon for operations such as grinding, polishing, buffing, lapping, etc., where controlled wear is desired. Lubrication is used in many operations to reduce friction and wear, although thick lubricant films can be detrimental with respect to final workpiece surface properties and dimensional tolerances that can be achieved.

Cross-References

- Abrasive Machining Processes
- Adhesive Contact of Inelastic Bodies
- Amontons Laws of Friction
- Chemical Conversion Coatings
- Chemistry of Rolling Lubricants
- Friction in Metalforming
- Friction Modeling for Machining
- Hydrodynamic Lubrication
- Hydroforming Tribology
- Lubricant Viscosity
- Lubrication Regimes
- Lubrication with a Newtonian Fluid
- Lubrication with a Non-Newtonian Fluid
- Reynolds Equation
- Surface Roughness
- Thermal Effect on EHL

References

- S. Kalpakjian, S.R. Schmid, *Manufacturing Processes for Engineering Materials*, 5th edn. (Addison-Wesley, Reading, 2008)
- S. Kalpakjian, S.R. Schmid, *Manufacturing Engineering and Technology*, 6th edn. (Prentice-Hall, Upper Saddle River, 2010)
- K. Laue, H. Stenger, *Extrusion* (American Society for Metals, Metals Park, 1981)
- S. Malkin, *Grinding Technology* (SME, Dearborn, 1989)
- P.K. Saha, *Extrusion Fundamentals* (in press)
- J. Schey, *Tribology in Metalworking* (American Society for Metals, Metals Park, 1984)
- J. Schey, *Introduction to Manufacturing Processes*, 3rd edn. (McGraw-Hill, New York, 1999)
- S.R. Schmid, W.R.D. Wilson, Tribology of manufacturing processes, in *CRC Handbook of Modern Tribology*, ed. by B. Bhushan (CRC Press, Boca Raton, 2001), pp. 1385–1411
- M.C. Shaw, *Metal Cutting Principles* (Clarendon, Oxford, 1984)
- M.C. Shaw, *Principles of Abrasive Processing* (Clarendon, Oxford, 1996)
- S. Sheu, W.R.D. Wilson, Mixed lubrication of strip rolling. *Tribol. Trans.* **37**, 483–493 (1994)
- O.J. Tassi (ed.), *Nonferrous Wire Handbook* (The Wire Association, Guilford, 1981)
- W.R.D. Wilson, J.A. Walowit, "An Isothermal Lubrication Theory for Strip Rolling with Front and Back Tension" *Tribol. Conv.* 1971 (Institution of Mechanical Engineers, London, 1972), pp. 164–172

Marine Engine Oils

TZE-CHI JAO¹, ANDRE VERHELST²

¹Department of R & D, Afton Chemical Corporation, Richmond, VA, USA

²Fomer Member of CIMAC, Texaco R&D, Ghent, Belgium

Synonyms

Low-speed two-stroke diesel cylinder oils; Marine lubricants; Medium-speed four-cycle trunk piston engine oils; System oils for large marine diesel engines

Definition

Marine engine oils comprise a series of lubricants used to maintain the proper function of low-speed and medium-speed marine engines, which are mainly employed in larger ships.

Scientific Fundamentals

Background

In 2009 more than 35,000 larger vessels ranging from 1000 dwt feeders to more than 500,000 dwt super oil tankers sailed across the oceans, transporting a wide variety of cargoes such as grains, coal, iron ore, cars, forest products,

manufactured goods, and crude oil. These ocean-going ships are generally powered by two types of diesel engines: low-speed two-stroke crosshead diesel engines and medium-speed four-stroke diesel engines.

It is noteworthy that, due to the vast advantage in economics of diesel engines over steam turbine engines, since the 1970s the latter have virtually been eliminated from ocean-going fleets. Low-speed two-stroke crosshead diesel engines are generally used as the main engine for cargo ships, container vessels, bulk carriers, and tankers. Medium-speed four-stroke diesel engines are typically used as main engines for cruise liners, ro-ro ferries, and smaller container vessels and as auxiliary engines for all types of larger ships.

The lubrication performance requirements for these two types of engines are quite different. The low-speed two-stroke crosshead diesel engines are lubricated by two types of marine lubricants – cylinder lubricants and system oils. Medium-speed four-stroke diesel engines are lubricated by only one type of marine lubricant – trunk piston engine oil. System oils and trunk piston engine oils are commonly referred to as marine crackcase oils. The lubricant composition and requirements of marine engine oils are substantially different from typical high-speed engine lubricants. The marine engines run mostly on heavy fuel (HFO) with high sulfur content versus low sulfur diesel fuel used in automotive diesels. The sump oil of marine diesels is seldom replaced for economical reasons. The oil is maintained in satisfactory condition by constant purification to remove water and contaminants from the oil.

To operate the large marine engines at an economical cost, it is of the greatest importance to minimize the fuel consumption. Therefore the fuel consumption is highly optimized, reaching 50% efficiency in modern marine diesel engines. This is the highest efficiency in the diesel industry. Secondly, the engines are designed to run on low-cost heavy fuel oil. This is a major economical advantage in comparison with high-speed diesel engines, which have lower efficiency and require more expensive high-quality diesel fuel. In fact the marine low-speed and medium-speed engines can run on a wide variety of fuels ranging from regular gas oil to very heavy fuel oil of 700 centistokes at 50°C.

There are substantial differences between marine engines and high-speed diesel engines in lubrication requirements. The differences in combustion pressure, fuel type, fuel combustion characteristics, fuel sulfur content, and others result in fundamentally different lubricants. The world-wide annual volume of the three types of marine lubricants is estimated to be about 2 million metric

tons or around 5% of the roughly 40 million metric tons of the total lubricants consumed in the world today.

Marine Diesel Engines

The basic designs of the two main diesel engine types used in marine ships are quite different from on-the-road and off-the-road high-speed diesel engines. Table 1 shows the differences in size and power output per cylinder of some of the largest low-speed and medium-speed marine engines versus an on-the-road high-speed engine. Additionally the marine low-speed and medium-speed diesel engines run mostly on heavy fuel oils (HFO). The high-speed diesel engines run on much more expensive high quality distilled diesel fuels.

Low-Speed Two-Stroke Crosshead Engines

Low-speed two-stroke diesel engines run at speeds between 60 and 130 rpm. Because of the low engine speed, the engine can be directly coupled to the propeller shaft without the use of a gearbox. Such design improves the mechanical efficiency. Figure 1 shows the typical layout of a low-speed diesel engine. In the crosshead engine design, for each cylinder, the piston rod and the connecting rod are linked to the reciprocating block called the crosshead, which slides up and down along the crosshead guide. An assembly of spring-loaded brass scraper rings referred to as a “stuffing box” prevents the physical mixing of the cylinder oil in the cylinder with the system oil in the crankcase compartment.

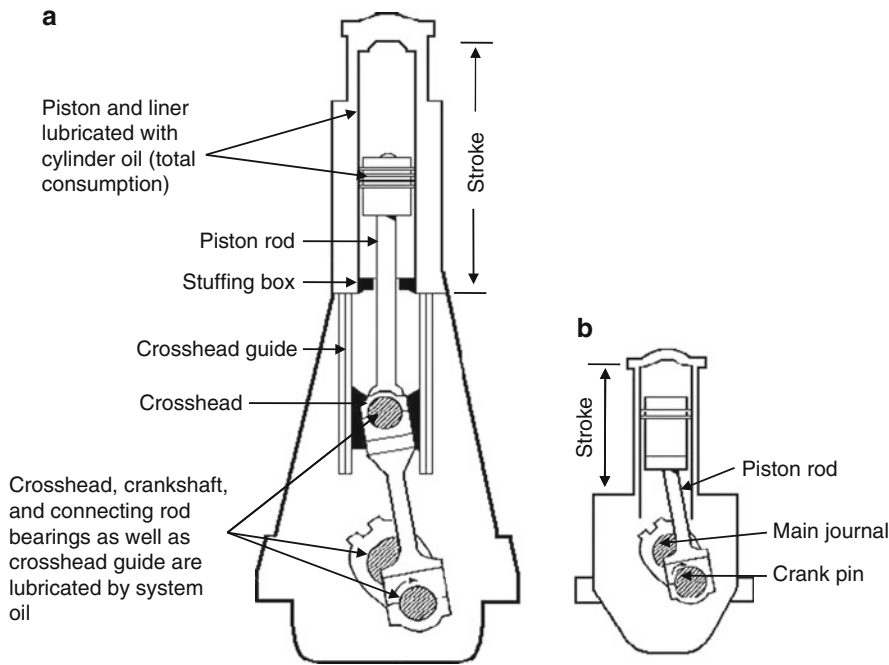
The cylinder oil is fed to the cylinder walls through a number of injection points called quills. Each cylinder can be fitted with 4 to 16 quills, depending upon the cylinder stroke and bore sizes (Carter et al. 1997). The quills are arranged circumferentially at either one or two levels and the oil is delivered to them by lubricating pumps. The system oil is used for forced lubrication of crosshead bearings, crankshaft bearings, and connecting rod bearings in the crankcase compartment. It also lubricates the crosshead guide. Furthermore, the system oil is used for piston cooling by means of circulating the system oil up the connecting rod into the piston under-crown space.

Trunk Piston Engines

Medium-speed diesel engines are often called trunk piston engines because of the piston skirt or trunk in a four-stroke engine transmits thrust to the cylinder liner in a similar manner as the crosshead slipper transmits thrust to the crosshead guide. In contrast to low-speed engines, the connecting rod is attached directly to the piston by a gudgeon pin. This design is similar to normal high-speed

Marine Engine Oils, Table 1 Comparison of slow-speed large crosshead and medium-speed engines with an on-the-road high-speed diesel engines

Crosshead engines			Trunk piston engines		High-speed engines
Engine type	MAN B&W K98MC7	Wartsila-Sulzer RTA96-C	MAN B&W 8L 58/64	Wartsila 46F	Mack MP7 US07
Bore (mm)	980	965	580	460	123
Stroke (mm)	2,660	2,489	640	580	152
Stroke: bore ratio	2.71	2.58	1.10	1.26	1.24
Engine speed (rpm)	97	102	428	600	1,200–2,000
Bmep (bar)	19.2	13.7–19.6	23	25.9	22.3
Output (bhp/cyl)	8,470	7,780	1,390	1,250	405



Marine Engine Oils, Fig. 1 Schematic drawings of: (a) crosshead engine and (b) trunk piston engine (Redrawn from Carter (1997))

diesel engines but considerably larger in dimension. A medium-speed four-stroke diesel engine runs between 250 and 850 rpm (Information can be found in the marinediesel.co.uk website provided by Warsash Maritime Academy of Southampton Solent University), while high-speed diesel engines typically run between 1,000 and 2,500 rpm. Because of the higher speed than a crosshead engine, medium-speed engines are coupled to the propeller shaft through a gear box when used as propulsion units. Medium-speed engines are also

extensively used as auxiliary engines for electrical power generation in all types of large ships. The differences in construction, size, speed, and output between low-speed engine and medium-speed engines are illustrated in Table 1 and Fig. 1.

Fuel Oils

Traditionally, low-speed two-stroke crosshead engines run on heavy fuel oil (HFO), while medium-speed four-stroke engines operate on a range of fuels from marine diesel oil

to heavy fuel oil. Heavy fuel oil is pure or nearly pure residual oil, which remains of the crude oil after processing and distillation. Marine diesel oil is usually a blend of distillate and heavy fuel oil. Up until 1982 there was no quality standard for marine fuel (Carter et al. 1997). The fuel was purchased based on viscosity and density. A British Standard, BS MA100, was introduced in 1982. An International Standard ISO 8217 was first published in 1987. The latest version of ISO 8217 was published in 2005. The parameters used in the International Standard for both marine distillate fuels and marine residual fuels include density at 15°C, viscosity at either 40°C or 50°C, water content, sulfur level, aluminum and silicon content, flash point, pour point in summer, pour point in winter, and cloud point. Calculated cetane index is included as a parameter in the standard for the marine distillate fuels.

The quality of marine fuel has a direct impact on engine operation. Fuel quality has deteriorated over time due to the increased use of cracking and visbreaking refining processes (Montaud et al. 1998). These refinery processes have been introduced to maximize the yield of premium products from crude oil. They produce heavier residues with poorer combustion properties and more impurities. Heavy fuel oils contain high sulfur levels, mostly in the range of 2.5–3.5%, with a maximum of 4.5%. High-sulfur fuel oil will produce correspondingly high concentrations of sulfur oxides during the combustion process. A portion of these sulfur oxides will be transformed to sulfuric acids. These acids are highly corrosive to the engine parts. High base number (BN) oils are required to neutralize the condensed acids.

Lubrication Environment

To some extent marine engine lubricants have similar properties as lubricants used in automotive engines, such as proper viscosity, high oxidation stability, high thermal stability, anti-wear properties, proper detergency, and dispersancy. The design factors to obtain these properties are extensively described in literature so no further discussion will be given here. However, marine lubricants have some unique requirements, which will be highlighted.

Marine engines are running on HFOs with high sulfur content. To prevent acidic corrosion in ring and liner areas, the cylinder oils and trunk piston engine oils must have sufficient high alkalinity reserve to neutralize the condensed sulfuric acids. To achieve the high alkalinity level in the oil, mixtures of large amounts of different highly overbased additives are used. The admixture of high concentrations of different highly overbased additives can cause some additive instability.

Therefore, component compatibility has to be taken into account and has to be tested. To maintain the oil in good condition without replacement, the crankcase oils of all larger marine engines are constantly purified to remove water and other contaminants. The crankcase oils must be able to release water and other oil contaminants in the purifier. Therefore good water separability is important.

In service trunk piston engine oils can become contaminated by HFO and combustion products of HFO, which contains highly aromatic asphaltenic types of materials. If these contaminants are not properly dispersed in the oil, they can coagulate and form deposits in the engine. Therefore marine-medium speed engines oils have to be especially designed for HFO compatibility. Another specific problem in marine engines is the occasional use of low sulfur-fuels in combination with high-BN lubricants, which can lead to high wear problem.

Some of these factors are discussed in more detail below:

Water Separability

Ships operate under humid conditions, therefore water condensation in crankcases is quite common. In addition, from time to time the oil may become contaminated by cooling water leakages or cooling water spills in the crankcase during maintenance of the engine. To keep the oil in good condition, marine engines are equipped with purifiers to remove water contamination and impurities from the oil. During purification, the oil is continuously in contact with a water phase. To assure proper water removal, it is important that no water emulsions are formed and that the water can properly be removed under centrifugal force. It is a very delicate balance between emulsibility and demulsibility of the oil. In the case of water contamination, no free water formation should occur in the engine, which could endanger the lubrication of the engine and at the same time water should be easily removed in a purifier. Water separability is very important for both crankcase oils – low speed engine system oils and medium speed trunk piston engine oils.

Corrosion Protection

In a marine environment sea water contamination occurs from time to time. Therefore, crankcase lubricants are expected to protect engine parts against rust and corrosion induced by sea water. In addition, the pistons and liners are exposed to high levels of sulfuric acids derived from the combustion of high sulfur-fuels. Therefore, low speed cylinder oils and the medium speed trunk piston engine oils must contain high amounts of alkaline additives to

neutralize the condensed acids in order to protect the liner and pistons from acidic corrosion.

Oil Filterability

Marine lubricants in service, like all engine oils, oxidize partially during use and are contaminated with combustion products. The main contamination products for engine oils are oxidation and nitration products, soot particles, partially combusted fuel, and calcium salts. The latter derive from the acid neutralization with calcium carbonate, present in the oil and from the oil ash originating from combusted oil. In marine engine oils, higher amounts of calcium salts will be present due to the high BN and the high ash content of the oils. All these contaminants can potentially coagulate to form larger particles, potentially causing filter blocking and engine deposits. In automotive crankcase lubricants the contaminants are dispersed in the oil by dispersants forming very fine particles, which do not affect the filtration and does not form engine deposits. To assure good condition of the oil and prevent excessive contamination levels, the oil is drained and replaced by new oil at certain time intervals.

In case of marine crankcase oils, changing the oil would be uneconomical because of the huge amounts of lubricant, up to 30,000 L in one sump. Therefore, the marine crankcase oils are continuously purified to remove the contaminants. This means that the contaminants must be dispersed to an average particle size small enough to be kept in suspension to avoid filter blockage and deposit formation in the engine but at the same time they must be sufficiently large to enable removal under centrifugal force. To achieve this delicate balance, a somewhat different additive balance is required. Marine lubricants contain higher amounts of detergents and lower amounts of dispersants than typical automotive engine oils.

Heavy Fuel Compatibility

Improved petroleum refining processes such as cracking and visbreaking allow extraction of higher amounts of light fractions out of the crude oil (Montaud et al. 1998). As a result, the residual fractions, which are used in marine residual fuels, contain a higher concentration of asphaltenes and contaminants. Asphaltenes are high molecular weight aromatic molecules, which are kept in colloidal suspension by their outer molecular structure. During thermal cracking asphaltenes lose part of their outer structure. This makes them less soluble in a paraffinic medium such as paraffinic fuels or lubricating oils. This characteristic can potentially lead to fuel instability; for example, highly cracked asphaltenes will not dissolve in fuels, which are too paraffinic. They will tend

to coagulate and form black asphaltic deposits. To assure fuel stability, refineries balance the degree of cracking and the aromaticity of the fuel.

Fuel pump leakage and incomplete combustion of the heavy fuel can lead to contamination of the lubricant with asphaltene type material. The cracked asphaltenes do not readily dissolve in lubricant. They tend to coagulate in the lubricant forming sticky asphaltic particles. These asphaltic particles have a high tendency to stick onto engine surfaces and can lead to engine crankcase blackening, oil filter blockages, oil scraper ring blockage resulting in increased oil consumption, and deposit formation in the piston cooling space. Formation of carbonaceous deposits in the piston cooling space, most typically on the piston undercrown, can act as an insulating layer resulting in higher piston temperatures as the heat cannot be removed by the lubricating oil. Temperatures above 450°C in the presence of sodium and vanadium, both being present in HFO, can lead to hot metal corrosion of the piston crown, which can ultimately result in hole formation in the piston crown. In this case, hot combustion gases can enter the crankcase. Potentially this can lead to a crankcase explosion. Coagulating asphaltenes can also form deposits on the purifier plates and in the purifier heaters etc. To prevent the above described phenomena, it is of great importance that marine trunk piston engine oils be designed to disperse asphaltenes.

Component Compatibility

Cylinder and trunk piston engine oils contain high concentrations of highly overbased detergents to combat acidic corrosion in the engines. For example, a typical low-speed engine cylinder oil has a BN of 70. This means a concentration of approximately 18–25% overbased additives. In order to achieve specific formulation benefits, the oil is often formulated with a mixture of different types of overbased detergents like calcium sulfonates and phenates. Mixing high amounts of these two overbased additives can result in colloidal instability and deposits formation. These deposits can potentially cause blockages of oil feed lines, filters, and lubricating quills. In the case of marine trunk piston engine oils, they are typically formulated around 20–40 BN. At lower concentrations the frequency of additive instability is much lower but it does occasionally occur in test formulations. It is important to select the correct additive admixture and to include sufficient stability testing during the evaluation process.

Impact of Low-Sulfur Fuel

Shipping is the most efficient type of transport in terms of fuel consumption per ton of goods transported over

a given distance. However, due to the high tonnages and long distances involved significant amounts of pollutants are still produced by the shipping industry. Until recently, there was no international legislation to reduce the sulfur level of marine fuels. In 2005, a cap of 4.5% sulfur in fuel was imposed globally by the International Maritime Organization (IMO) – a United Nations body. For environmental reasons, a cap of 1.5% sulfur in fuel was imposed in the Baltic Sea in 2006 and a cap of 1.5% sulfur in fuel was imposed in the North Sea in 2007 (Vrolijk et al. 2008).

The sulfur content in HFO varies depending on crude oil origin and the refinery processes. The impact of the legislation and the natural variation of sulfur content in the HFO mean that ships sometimes operate with both high- and low-sulfur fuels. For example, a ship runs with a typical 3% sulfur fuel when it sails across the ocean for economic reason then switches to a 1.5% sulfur fuel when it enters SOx Emission Control Areas (SECA), which includes Baltic Sea, North Sea, and English Channel.

It has been observed that the use of low-sulfur fuels in combination with high-BN lubricants increases the risk for scuffing on the cylinder liners. Investigations indicate that the scuffing is induced by the rubbing of piston crown deposits against the liner. If low sulfur fuel is used in combination with high BN oils, higher amounts of piston crown deposits are formed. This is due to excessive alkalinity supply in the form of calcium carbonate, which does not react away due to low concentrations of sulfuric acids in the combustion chamber. Bringing the supply of the alkalinity in balance with the fuel sulfur content resolves the problem in most cases. In practice, problems with low sulfur fuel in low-speed engines can be resolved by applying a low BN cylinder oil, e.g., 40 or 50 BN or by reducing the cylinder oil feed rate still maintaining it within the engine manufacturer recommendations.

To prevent excessive alkalinity supply in trunk piston engines, it is recommended to maintain the BN of the sump oil relatively close to the OEM recommended minimum level by adapting the BN of the topping up oil if required. It might be necessary to use lower or higher BN oils to maintain the recommended level. This will depend on the sulfur content of the fuel and the oil consumption of the engine.

Key Applications

There are three types of marine lubricants being used presently. They are cylinder oil and system oil used in low-speed two-stroke crosshead engines and trunk piston engine oil for medium-speed four-stroke engines. Since marine engines normally run in steady temperature conditions, only mono-grade engine oils are required.

Cylinder Oils

Cylinder oils are used in low-speed two-stroke crosshead engines. They are directly injected in the cylinder.

The lubricant has a very short life in actual operation. A typical marine cylinder lubricant meets SAE 50 viscosity grade. A high viscosity grade of SAE 50 is required to provide a strong oil film between the cylinder liner and the piston rings. An SAE 50 will still provide sufficient viscosity at high pressures, high liner temperatures of around 230 °C and low speeds. The lubricant is typically formulated with an alkalinity reserve of 70 BN for engines operated with high-sulfur heavy fuel oil. For low-sulfur fuels a 40–50 TBN cylinder oil is often employed.

A high alkalinity reserve is required to neutralize the condensed acids in the upper part of the liner and ring belt area. In addition, the lubricant has to be sufficiently oxidation and thermal stable to prevent formation of deposits by thermal decomposition and oxidation and nitration of the lubricant. The oil has to disperse the different contaminants such as soot particles, partially burned fuel residues, calcium salts from the neutralization process, and oil ash from lubricant combustion to maintain satisfactory engine cleanliness.

To achieve satisfactory wear performance in low-speed engines, all factors have to be optimized, including the engine design, the operations, and the lubricant quality, in order to minimize corrosive wear, abrasive wear, and mechanical wear in the engine.

Corrosive wear is combated by a high alkalinity reserve in the lubricant by an optimized oil feed rate to neutralize the condensed acids effectively and by optimized liner temperatures to prevent excessive acid condensation. Abrasive wear can be caused by abrasive particles from the fuel, like catalytic fines originating from the fuel catalytic cracking process or rust particles from the fuel transport. To prevent fuel particle abrasion, it is essential that the heavy fuel oil is properly purified to eliminate abrasive particles from the fuel as much as possible. Abrasive wear can also be caused by hard calcium deposits on the piston crown, which rub against the liner surface.

The formation of piston crown deposits can be minimized by optimizing the thermal stability of the lubricant, by optimizing piston cooling reducing the piston crown temperatures, and by optimizing the oil feed rate and the alkalinity of the lubricant to prevent excessive amounts of calcium deposits.

The piston crown deposits can also be mechanically removed by the introduction of an anti-polishing ring. An anti-polishing ring is mounted in the liner and is scraping off the deposits from the piston crown land, preventing that deposits touch the liner.

Thus, it is important to apply very highly thermally stable cylinder oils and to optimize the oil feed rate to prevent excess BN supply leading to formation of abrasive deposits. In the case of using low-sulfur fuel it is recommended to use a lower BN cylinder oil or to reduce the oil feed rate within OEM guidelines.

Mechanical or adhesive wear is in most cases induced by the same mechanism as abrasive wear. It is a result of liner polishing by piston crown deposits, finally resulting in scuffing. Mechanical wear can be minimized by the same measurements that prevent abrasive wear.

The cylinder of a low-speed two-stroke crosshead engine has a large surface area. Cylinder oil application is attained by timed injection through 4–16 quills spaced equidistantly around and near the top of the liner. In order to properly lubricate the entire surface areas of the liner and piston, cylinder oil must possess a sufficiently good spreadability. This oil property may be determined in the laboratory by dropping small quantities of oil onto a flat metal plate held at an elevated temperature of about 200°C and measuring the area covered by the oil. Base oils alone have good spreadability, but this can be reduced significantly by high concentrations of overbased additives incorporated into the cylinder oil. This can be alleviated by blending in a suitable amount of certain surface-active additives.

During development of developing a cylinder lubricant, oils are screened in a series of bench tests to evaluate acid neutralization rate, colloidal stability, spreadability, water stability, oxidation and thermal stability, anti-wear performance, and deposit formation at high temperatures. The next step is to evaluate the candidates in engine tests. For evaluation of cylinder lubricants some laboratories use a two-stroke diesel Bolnes engine, running on highly cracked heavy fuel. The final proof of a cylinder oil performance is established by field testing in a series of ship's engines. At the end of the test, the engines are assessed for liner wear, ring wear, and cleanliness. Engine manufacturer approvals are obtained based on proven field performance. After introduction, the oil performance is continuously monitored in a high number of different types of engines.

System Oils

The second lubricant required in a low-speed two-stroke crosshead engine is the system oil. The main function of the system oil is to lubricate three types of bearings and the crosshead guide in the crankcase (Mang et al. 2007). In addition, the system oil is used for piston cooling and for lubricating the bearings of integrated turbo chargers. To

minimize the number of lubricants on board, the system oil is frequently used for lubrication of other equipment such as the stern tube, some deck equipment and low-load gear transmissions.

The oil viscosity needs to meet the SAE 30 viscosity specification. It needs to provide good bearing performance, which means apart from proper viscosity, the oil has to be clean and provide rust and oxidation protection. To prevent deposit formation in the piston cooling space, the system oil needs to be sufficiently thermal and oxidation stable. The oil should have a mild alkaline reserve to neutralize condensing acids. Although there is no direct contact with the combustion chamber, occasional contact with the acidic combustion gases from combusting high-sulfur fuel cannot be fully excluded. A typical BN of a system oil is around 5 BN. Even though system oil is separated from cylinder oil by a stuffing box, some contamination with used cylinder oil occurs on a regular basis. The contamination causes an increase of viscosity and an increase of BN by admixture of used cylinder oil with a high residual BN. Other contaminants originating from used cylinder oil should be easily removed by the purifier. Furthermore, system oils require good demulsibility and should provide corrosion protection against salt water contamination. The latter properties are discussed in more detail below.

Demulsibility

Demulsibility is an important property for system oils because system oils are occasionally contaminated with water. The water can originate from different sources such as leakage from the piston cooling system in some older engines, where water cooling is applied. It can also originate from water condensation in the crankcase or from cooling system leakages or from water ingress during maintenance or from an improperly functioning purifier. Since the amount of system oil can be substantial, up to 30 t, for economical reasons it is of great importance that water contamination from any source can be easily removed. Ideally, water can be properly removed under centrifugal force and no free water is formed during dynamic conditions to protect the bearings.

Salt Water Corrosion Protection

Marine engines operate in a humid and salty environment. The crankcase oil can be contaminated with salt water. Particularly if the system oil is used in the stern tube, there is a direct contact with seawater. Stern tube

oils are from time to time contaminated with seawater. To protect the engine and the stern, it is required that the system oil has the ability to protect against salt water corrosion.

Purification Properties

The purifier maintains the oil in good condition by removing water and particle contaminations. The latter could originate from used cylinder oil contamination. To assure good condition of the system oil, it is important that, apart from good water shedding, the oil contains a mild detergency to keep the combustion particles contamination in suspension to prevent deposit formation. Particle size should be sufficiently large to enable removal in the purifier. In practice, the system oils contain a mild detergency and no dispersants to keep the engine clean without upsetting the purification operations.

Load Carrying Capacity

If the low-speed crosshead engine is equipped with a power take-off, a system oil with higher load carrying capacity has to be employed. The load carrying capacity is evaluated by the FZG test. The desired load carrying capacity is commonly achieved by the use of a small amount of zinc dialkyl dithiophosphate (ZDDP).

Product Development

During development, the system oil is tested in a series of bench tests to evaluate oxidation and thermal stability, anti-rust performance, and anti-wear performance. The deposits formation tendency at high temperatures is often evaluated in the high-temperature panel cooker test wherein oil is splashed onto a hot metal plate and the produced deposits are quantified. Water stability, water separation, and purification properties after contamination are normally evaluated in a small purifier. The next step is to evaluate the bearing protection performance in an engine test under real engine conditions. Before commercialization, the oil will be field tested in a series of real marine low-speed engines. The condition of the oil is intensively monitored by used oil analyses during the field testing. At the end of the field test the engines are assessed for bearing condition and engine cleanliness. Engine manufacturer approvals are obtained based on proven field performance.

Oil Analyses

To assure good oil condition in use, the oil is monitored by regular lubricant analyses.

Trunk Piston Engine Oils

Typical viscosity grades for trunk piston engine oils are SAE 30 and SAE 40 grades. The applied viscosity grade is recommended by the engine manufacturers. The alkaline level varies from 12 to 50 BN depending on the sulfur content of the fuel and the oil consumption of the engine. In practice, a 40 BN oil is most often recommended for HFO operations. Like other crankcase engine oils, the main role of marine trunk piston engine oils is to lubricate the engine, to control piston deposits, to protect bearings, and to provide good anti-wear performance. Marine medium-speed oils have a number of additional requirements.

The oils should provide sufficient alkalinity to neutralize the condensed sulfuric acids. The amount of alkalinity supplied to the liner and the ring belt area depends directly on the BN of the used oil in the sump. The BN of the sump oil is an equilibrium value depending on the BN consumption, which depends on the sulfur content of the fuel and on the added BN, which in turn depends on the oil consumption. To assure proper neutralization of condensed acids in medium speed engines running on HFO, most manufactures recommend maintaining a minimum BN level in the crankcase between 18 and 20 BN. In more modern engines, the oil consumption has been reduced from 0.4 g/kWh to less than 0.15 g/kWh. In the case of low oil consumption in combination with high-sulfur fuels, the BN equilibrium value will drop easily below the recommended alkalinity level. To prevent costly oil drains, very high BN oils between 55 and 60 BN are used to top up the system to maintain a satisfactory minimum BN level.

Marine trunk piston engine oils should provide good filterability and enable proper purification. This means the lubricant should be formulated in such a way that the contaminants like soot particles and calcium salts are sufficiently dispersed to prevent deposit formation and/or filter blocking but at the same time that the size of these particles is sufficiently large to enable removal by centrifugal force in the purifier. The oils should also have good balanced water shedding properties. This means no free water formation during dynamic conditions, which can endanger the bearings, and proper water separation properties under centrifugal force to enable water removal in the purifier.

HFO Compatibility

Another important property of marine medium speed engines lubricants is heavy fuel oil compatibility. Medium-speed engine oils are potentially contaminated with HFO and combustion products of HFO. The fuel contains asphaltenes, which can form sticky floating

particles in the lubricant. These particles cause serious problems in the engine. They can stick to engine parts causing engine blackening, form deposits in the ring belt area, in the cooling space of the piston and in purifiers and purifier heaters. Additionally, the asphaltic particles can cause increased bearing wear by particle erosion in soft bearing types. The increased bearing wear is a result of particle erosion, excessive imbedding of particles in the bearing surface, and bearing corrosion from particles originating from the combustion chamber containing absorbed acids.

The excessive ring belt deposits can lead to increased oil consumption by blocking the holes in oil scraper rings. Deposits in the piston cooling space form an insulation layer reducing piston cooling, resulting in higher piston temperatures. Above 450°C there is a risk for high-temperature corrosion resulting in material loss on top of the piston. This can lead to early replacement of piston heads and, in the worst case, to hole formation in the pistons.

All above illustrates the great importance of HFO compatibility of marine trunk piston engine oils. HFO compatibility means that oil must have the capability to disperse highly cracked asphaltenes originating from HFO contamination and incompletely combusted HFO to prevent deposit formation. Different tests have been developed to determine the asphaltene content in an oil and to evaluate the dispersion capability of a trunk piston engine oil.

The capability of an oil to disperse asphaltenes is evaluated by studying mixtures of lubricant and highly cracked HFO. The oil quality is evaluated by direct measurement of size and amount of particles formed in the oil or by determining the amount of deposits in a filtration test or by measuring the deposit formation tendency in deposit test.

Product Development

During the development stage the oils are evaluated in a number of bench tests, for oxidation stability, thermal stability, fuel compatibility, and deposit formation at high temperature. The oils are evaluated in small marine lubricant purifiers to screen for water separability and the additive stability against water. Used lubricants from engine testing are employed to measure the speed of removal of contaminants in a laboratory purifier. The candidate oils may also be tested in a number of standard laboratory engine tests to evaluate general diesel performance. The oils are evaluated for bearing wear, piston cleanliness, engine cleanliness, anti-bore polishing performance, under-crown deposits formation, and ring and liner wear. In addition, the oils may be evaluated in a laboratory test engine, which is adapted to run on highly

cracked HFO. Sometimes the tests are run with contaminated lubricant with 2–10% heavy fuel oil to create severe fuel contamination conditions (Vrolijk et al. 2008).

There is no standard HFO engine test in the industry to assess diesel detergency and anti-wear performance of marine trunk piston engine oils. Trunk piston engine oil suppliers must develop their own in-house tests (Saile and Maycock 1977). After full laboratory evaluation the candidate lubricant is field tested in a series of real marine medium-speed engines from different manufacturers. The testing occurs in full cooperation with the engine manufacturers. During the tests the oil is monitored by used oil analyses. At the end of the field tests, the engines are evaluated for general engine cleanliness, piston cleanliness, under-crown deposits, ring and liner wear, and bearing condition. Engine manufacturer approvals are obtained based on proven field performance. In service, the oil performance is constantly monitored in a number of engines to assure good performance.

Cross-References

- [Additive Chemistry Testing Methods](#)
- [Corrosive Wear](#)
- [Detergents](#)
- [Engine Lubricants](#)
- [High Temperature Lubricants](#)
- [Lubricant Viscosity](#)
- [Lubrication Regimes](#)
- [Lubrication with Emulsions](#)
- [Used Oil Analysis](#)

References

- B.H. Carter, Marine lubricants, in *Chemistry and Technology of Lubricants*, ed. by R.M. Mortier, S.T. Orszulik, 2nd edition (Blackie Academic and Professional, London, 1997), pp 237–254.
- T. Mang, W. Dresel, Marine diesel engine oils, in *Lubricants and Lubrication*, 2nd edition (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007), pp 227–229.
- A. Montaud, W.P. Fabrick, D.J.E. Vrolijk, K.C. Lim, A. Dunn, Contamination of marine crankcase lubricants by raw fuel: consequences, and methods of detection. *J. Inst. Energy*, Vol. 71, March, pp 2–11 (1998).
- J.A. Saile, J.M. Maycock, Laboratory evaluation of marine oils for medium speed engines burning distillate and nondistillate fuels. *J. Amer. Soc. Lubri. Eng.* February, pp 19–35 (1997)
- D.J.E Vrolijk, Challenges and solutions for marine engine oil technology, in *Proceedings of the 14th Annual Fuels and Lubes Asia Conference*, Seoul, Korea, March 5–7, 2008, pp 2–15.

Marine Lubricants

- [Marine Engine Oils](#)

Mass Transfer in Polymer Friction

► Polymer Friction Transfer (FT)

Materials for Mechanical Seals

MICHAEL HUEBNER¹, LIONEL A. YOUNG²

¹Principal Engineer, Flowserve Corporation, Deer Park, TX, USA

²Advanced Technology, Flowserve Corporation, Temecula, CA, USA

Synonyms

Aluminum oxide; Ceramics; Mechanical seal face materials; Metallized bound carbon; Reaction bonded silicon carbide; Resin bound carbon; Self-sintered silicon carbide; Silicon carbide; Silicon nitride; Tungsten carbide for seals

Definition

Materials used for the manufacture of mechanical seal faces. These materials must meet the tribological demands unique to mechanical seals. Industry has evolved primarily to the use of mechanical carbons and ceramics for these components. A mechanical seal (Fig. 1) is by its very nature a collection of components that are designed to control leakage of a fluid through a sealing interface. These components serve a wide variety of functions and correspondingly require a wide range of material properties. Many of these properties, such as mechanical strength, thermal conductivity, or corrosion resistance, are similar to considerations made for many other devices. Some properties, however, are unique to mechanical seals, such as coefficient of friction and wear. These considerations also vary significantly depending upon the specific component in the seal. All the materials must work together, however, to allow the seal to function properly.

While there are virtually an unlimited number of materials that could be used to construct a mechanical seal, the sealing industry has narrowed the selection to a relatively small number. These materials were chosen due to their properties, global availability, ease of manufacturing, and costs. Because these materials have had such a widespread use in seals, there is a very large installed base of successful applications that an end user or seal OEM can reference when selecting materials. While some of the material selections have been captured



Materials for Mechanical Seals, Fig. 1 Image of typical mechanical seal faces

in standards (e.g., API 682/ISO 21049), most of the selections have been made by the seal OEMs to meet the specific requirements for their seals in their intended applications.

Scientific Fundamentals

Mechanical seal face materials must have the following attributes:

1. Favorable tribological properties for operation in boundary layer and mixed film regimes on a wide variety of fluids
2. Mechanical properties that allow for consistent control of mechanical deformations
3. Sufficient thermal properties to allow for removal of heat and controlled temperature deformations

Seal Faces

The sealing interface is rightfully considered to be the heart of the mechanical seal. This sliding surface between the rotating and stationary faces provides a significant challenge for any material. Unlike other sliding surfaces such as bearings, a mechanical seal face may be operated over a wide range of surface speeds with a variety of face loadings while being lubricated by virtually any liquid seen in a centrifugal pump. Because of this range of applications, there is no single seal face material combination that can work successfully in every application. There are,

however, some general guidelines that are used in seal face material selection.

Seal faces are commonly operated so that there is some mechanical contact while there is relative motion between the faces. During start-ups, shutdowns, or upsets in a centrifugal pump, the faces may have significant contact with little lubrication. For this reason, seal faces materials are commonly selected so that one face is a hard material and one is a self-lubricating soft material. The hard material, generally a ceramic, acts as a non-wearing, reference surface. The soft material, generally a mechanical carbon, acts as a wearing, self-lubricating surface. In some applications, especially if small, hard particles are present in the sealed fluid, two hard faces may be used to minimize abrasion or damage of the sealing interface. In this case, other seal design factors such as a reduction in face loading or the addition of hydrodynamic face features may be required to ensure that an adequate fluid film is present.

A mechanical seal face does not act in isolation. It is always in contact with its mating seal face. This tribological pairing helps define the performance of the seals. Certain combinations of face materials have proven to be more successful than others and these have become default standards throughout the seal industry. An example of this is reaction bonded silicon carbide running against resin impregnated carbon. These mating pairs are generally characterized by low friction and tolerance of heavy contact. The concept of mating pairs between seal face has proven to be useful in seal design and application.

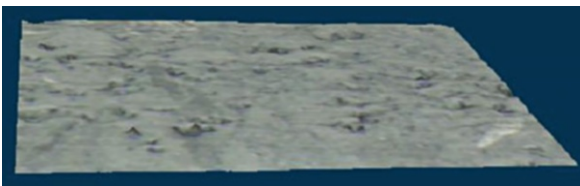
Requirements

The selection of materials for mechanical seal faces is driven by the tribological interaction between the seal faces as well as the characteristics of the fluid film and fluid properties of the sealed fluid. Designers of most lubricated equipment intentionally create conditions where the mating surfaces are prevented from contacting to minimize surface wear. Mechanical seal faces, however, are frequently designed so that the fluid film is diminished to the point where there is significant contact between surfaces asperities. This is done to minimize leakage between the seal faces. Operation in this mixed lubrication regime places higher demands on seal face materials. Although the materials are lubricated by the sealed fluid, characteristics such as the effective coefficient of friction and the predicted wear rates fall between dry running and full fluid film operation and can vary significantly depending upon the specific application.

Mechanical seal faces are always assumed smooth and flat to help prevent liquid leakage between the faces. In practice, seal faces are manufactured to create a surface



Materials for Mechanical Seals, Fig. 2 Carbon seal face as lapped, 300X

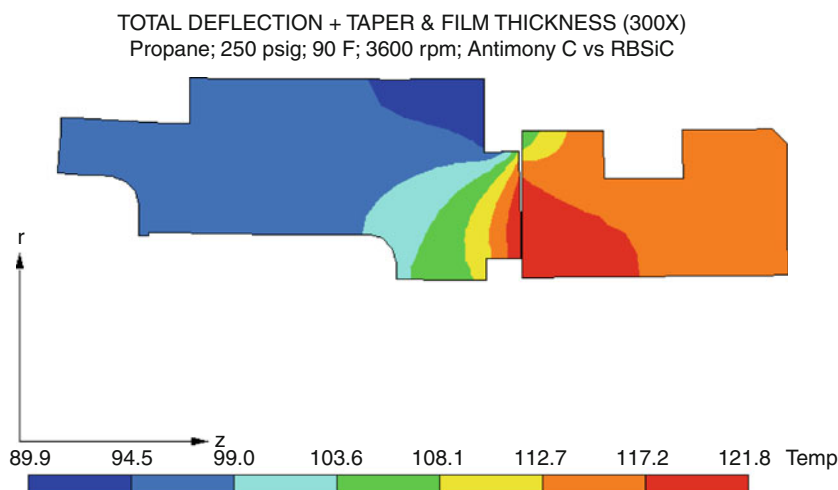


Materials for Mechanical Seals, Fig. 3 Carbon seal face after operation, 300X

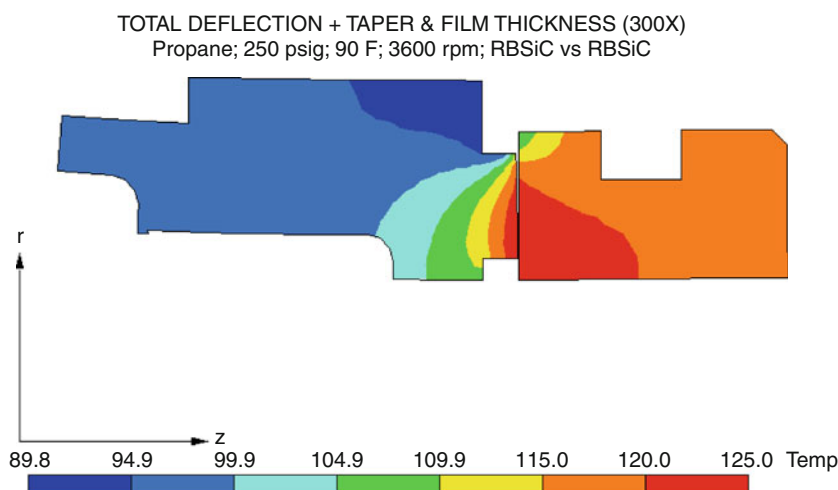
roughness, which improves the tribological condition between the faces. The surface roughness allows the faces to contact on the asperities to provide mechanical load support but also allows for the fluid to ingress the sealing interface to help provide lubrication and hydrostatic load support. In practice, the surface roughness is controlled by the lapping process (e.g., media, particle size, and lapping conditions), which seal OEMs have developed for their specific materials and design parameters. A typical view of a lapped surface is shown in Fig. 2. After operation, even seals faces that are operating properly will experience some wear as the mating asperities contact and are removed (Fig. 3). This results in a condition commonly called a “wear track,” which is characterized by a polished surface with residual roughness left in the depressions and is considered normal for most seal applications.

Materials have been created that can create their own surface features and alter the sealing interface. Voids on the surface or non-homogeneous materials can create small pores in the sealing interface and provide pockets of lubrication and even hydrodynamic support for the faces. This technique, while functional, has seen limited use and is generally limited to seal applications where an adequate fluid film cannot be created by conventional means.

By definition, a mechanical seal will be operated with a pressure differential across the seal faces. This pressure can be assumed to act uniformly over the surfaces wetted by the higher pressure fluid. The mechanical seal must contain these pressures without structural damage.



Materials for Mechanical Seals, Fig. 4 Resin impregnated carbon vs RBSiC



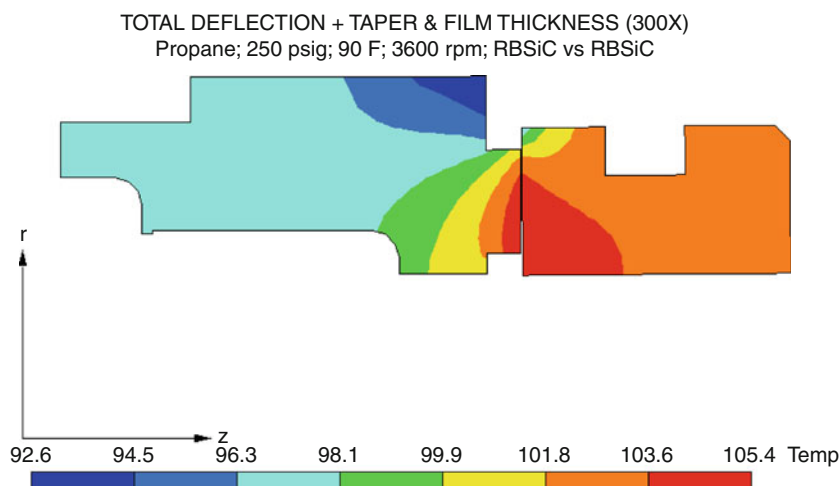
Materials for Mechanical Seals, Fig. 5 Antimony impregnated carbon vs RBSiC

Many seal face materials have excellent mechanical properties in compression but suffer from significantly lower properties in tension and can fracture rapidly in a brittle failure mode. The material selection can therefore have a significant impact on the design of specific components and the operation of the seal. Face materials must also resist deformations from this pressure differential. These deformations will affect the formation and shape of the fluid film. While a higher modulus is desirable, it is more important that it is consistent so the seal designer can compensate for these mechanical properties in the seal design.

Mechanical seal faces generate heat in the contacting interface between the faces due to mechanical friction and

fluid shear. In a successful seal application this heat must be managed to control thermal deformations and degradation of the fluid film. This implies that the seal face materials must be capable of conducting this heat to a region of the seal where it can be conducted away by convection into the process fluids. Face materials with high thermal conductivity will exhibit lower thermal distortion and allow for easier control of face deformations.

Due to the complex interactions between the seal face materials, operating conditions, and fluid properties, the selection of seal face materials is a complex engineering decision. This also makes it risky to arbitrarily change materials without a thorough analysis. [Figures 4, 5, 6](#)



Materials for Mechanical Seals, Fig. 6 RBSiC vs RBSiC

show an analysis of the same seal operating under the same conditions with three different face material combinations. This illustrates the interactions between some of the factors previously discussed.

Soft Faces

In mechanical seals, the term “soft face” is used to describe the seal face that will normally wear in operation. As stated earlier, it generally provides some self-lubricating properties that will allow the seal to tolerate hard contact or poor lubricating process fluids. Mechanical grades of carbon are almost exclusively used for these faces. The most common grades of carbon used for mechanical seals are:

- Resin bound carbon
- Metallized bound carbon
- Dry-running grades
- Acid grade

Carbon is one of the most abundant elements on earth. It is the basis for all organic products and processes. It is also an interesting material because it takes on forms from amorphous carbon to graphite to diamonds to fullerenes. Carbon is inert, stable, and can be self-lubricating. It is used in products as varied as pigments, carbon black in rubbers, and electrical brushes. It can take the form of soft carbon graphite powder or hard friction pads.

Mechanical carbons used in seal faces are a mixture of amorphous carbon and graphite. The percentages of each help determine the physical properties on the final grade. In addition to carbon, other elements and compounds are present that affect the properties of the grade of carbon.

Some of these are impurities from the original sources of carbon; others are specifically added to improve some aspect of the carbon's performance. These include additives to control film formation such as silica, silicon carbide, molybdenum disulphide, or lithium fluoride.

Mechanical carbons are manufactured by blending together types of amorphous carbon (i.e., lampblack, charcoal, or coke) and carbon graphite with a carbonaceous binder (i.e., pitch or resins). The source of the raw materials not only helps determine the physical properties of the carbon but also the type and amount of impurities. Other additives will be blended in depending upon the specific grade. The mixture is then pressed into shapes and heated in an inert environment. At a high temperature, most of the binder decomposes into carbon while a small amount volatilizes and leaves the part. This leaves the carbon porous and soft.

The carbon is then placed into a vacuum chamber to remove any air from the porous carbon. While under a vacuum, the carbon is immersed into liquid impregnant. The vessel is then pressurized to drive the impregnant into the carbon. This effectively fills the pores of the carbon, making it impermeable to fluids and greatly increasing the strength of the final material.

The selection of impregnants is a critical factor in determining the properties of the final material. Impregnants include various thermoset and thermoplastics resins, metals, and salts. The most common impregnants for mechanical seal faces are thermoset resins and antimony metal. While a carbon manufacturer may produce hundreds of grades of carbons, the mechanical seal industry has standardized on only a dozen or so grades that are

used for the majority of seal applications. With the consolidation of companies in the carbon manufacturing business, the list is fairly small.

Resin Bound Carbon

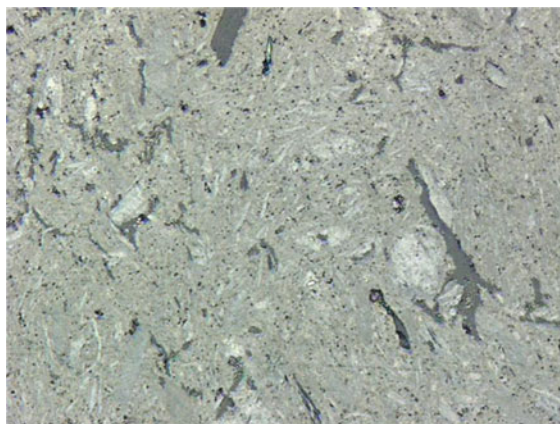
As the name implies, resin bound carbon (also referred to as resin impregnated carbon) is a mixture of amorphous carbon and graphite that has been impregnated with a thermoset resin (Fig. 7). This is by far the most common type of carbon for mechanical seals used in industry today. While there is a wide variety of specific formulations or grades available, most resin impregnated carbons are capable of operating in wide range of chemicals from strong bases to strong acids. They possess good frictional properties and an adequate modulus to help control pressure distortions.

The resin impregnant in carbon materials adds significantly to the mechanical properties of the carbon. In general, the higher the percentage of resin in the material the less desirable it is for mechanical seal faces. The most common resin materials generally have a melting point around 260°C (500°F). This is below the operating temperature window for many mechanical seals. In practice, high quality carbon grades have very low resin content; they work excellently even at temperatures well above this range and are the standard selection for high temperature seal applications.

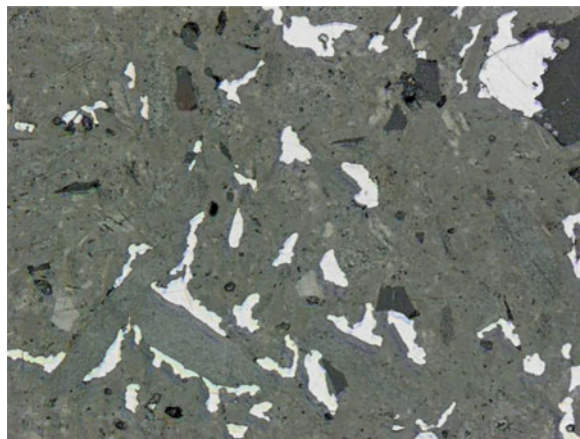
Antimony Bound Carbon

Metallized carbons are available with a variety of metal impregnants including copper, bronze, lead, and antimony. Of these, antimony has proven to be the most successful in seal applications (Fig. 8). This material is often referred to as antimony bound carbon or antimony impregnated carbon. The addition of antimony has several beneficial effects that can improve seal performance. First, the addition of a metal impregnant increases the strength and modulus of the material. This is beneficial for high pressure applications when a stronger and stiffer material is needed. The metal phase in the material increases thermal conductivity to control thermal distortion. Finally, antimony impregnated carbons are more resistant to blistering in high viscosity fluids or light hydrocarbons. This has made it the standard grade for many refinery applications.

These benefits come at a price. The chemical compatibility of an antimony carbon is limited by the antimony metal. Carbon manufacturers and seal OEMs can give guidance on selection of antimony impregnated carbons for specific applications.



Materials for Mechanical Seals, Fig. 7 Resin impregnated carbon, 100X



Materials for Mechanical Seals, Fig. 8 Antimony impregnated carbon, 100X

Dry-Running Carbons

Some seal applications are specifically designed to run without a fluid between the seal faces. In some cases, the seal faces are designed to lift off and not contact in operation. In other cases, the faces are designed to run with contact and no lubricating liquid. These dry-running, contacting seal designs create serious demands on the carbon and mating face of the seal. The carbon material must be self-lubricating and have a low coefficient of friction and yet be strong and durable enough to run for years without wearing out. The design parameters of mechanical seals are also adjusted to minimize face contact loading, which results in lower heat generation and lower wear.

Special grades of carbon have been designed for dry-running conditions. These grades are generally softer than other mechanical seal grade carbons. This is primarily due to the high graphite content in the mixture. Other additives such as salts (i.e., barium fluoride, lithium fluoride) can be added to improve performance by aiding in film formation between the faces. In most cases, a small amount of moisture in the seal environment greatly improves the performance of these grades.

Acid Grades

As stated earlier, elemental carbon is inert to attack from most chemicals. It was also stated earlier that most commercial grades of carbon are mixed from a variety of carbon sources. These sources differ greatly in the type and amount of impurities that are introduced into the mixture. The impurities are generally not resistant to aggressive chemicals like strong acids. This can lead to leaching and weakening of the component. Processing steps such as graphitizing can also affect chemical compatibility. To improve the carbon's performance, very pure sources of carbon are used to produce the most chemically resistant grades. These grades can compromise other aspects of the carbon properties and seal performance so they are generally reserved only for aggressive chemical applications.

Hard Faces

Hard faces can be described as seal faces that are not intended to wear in operation and are intended to provide a smooth, flat contact surface for the mating face. Historically, these surfaces have included materials such as hardened steels, metallic overlays, ceramic overlays, and ceramics. While there are still a variety of materials used across the industry, the majority of mechanical seals are currently supplied in solid ceramic.

Ceramics in a broad sense encompass a class of materials defined as non-metallic, non-organic materials that usually require high temperature processing. In a general sense, this can refer to everything from pottery and china to carbides and oxides. In an engineering sense, the term generally refers to a class of materials that are characterized by their high hardness, high stiffness, low thermal expansion, and good wear resistance. For mechanical seals, these include silicon carbide, silicon nitride, tungsten carbide, and alumina oxide. While these materials share similar characteristics in properties such as hardness and modulus, this range of ceramics varies significantly in other properties such as thermal conductivity, thermal shock resistance, strength, impact resistance, and chemical compatibility. Manufacturing techniques are



Materials for Mechanical Seals, Fig. 9 Raw manufactured SiC material

also a consideration in the selection since they can affect the allowable shapes and cost of the faces.

Silicon Carbide

In the simplest sense, silicon carbide (SiC –Fig. 9) consists of one atom of silicon bonded to one atom of carbon. This results in a tenacious bond that is stable over a wide range of temperatures and chemical environments. It also has other desirable properties such as a high hardness and a high modulus. Unfortunately, it is also a material that is difficult to manufacture in shapes suitable for component design. It almost never exists in nature. Although it was first identified in 1824 it remained somewhat of a curiosity until 1892 when a practical manufacturing process, the Acheson process, became commercially available. This process, a carbon reduction of silica, results in large bluish-black crystals that must be powdered and further processed into useful shapes. Other manufacturing processes to create SiC such as carbon conversion and chemical vapor deposition (CVD) have a more limited use in mechanical seal faces.

Reaction Bonded Silicon Carbide

Reaction bonded silicon carbide (RBSiC), as its name implies, is a material formed by bonding SiC particles to each other in a reaction process (Fig. 10). SiC particles and carbon are mixed together with a binding agent and pressed into the desired shape. When manufacturing seal faces, it is common for SiC manufacturers to isostatically or uniaxially press the material into a tube shape since this minimizes the secondary machining and grinding required to manufacture the part. In some cases, the material is further machined in the green state to final dimensions or to add features that would be most costly to add after firing.

The green state material is placed into a furnace with an inert atmosphere and exposed to molten silicon metal. The silicon metal wicks its way throughout the material reacting with the free carbon to form additional SiC. The reaction literally bonds the original SiC particles together by forming additional SiC material, which acts as a “glue.” Done correctly, this creates a fully connected matrix of SiC material. As a result of the manufacturing process, all of the spaces in the material are filled with silicon metal. In current grades, 8–12% of the final composition is free silicon. There is negligible shrinkage (<1%) during this process.

While the silicon metal does not significantly affect most of the physical and thermal properties of the material, it does limit the chemical resistance of the material. Anything that can chemically attack elemental silicon can attack the interconnected passages of silicon in RBSiC. This can weaken the material and can cause seal failures. The most common chemicals that will attack RBSiC are caustics (and other high pH chemicals) and strong acids. RBSiC should not be used with these applications.

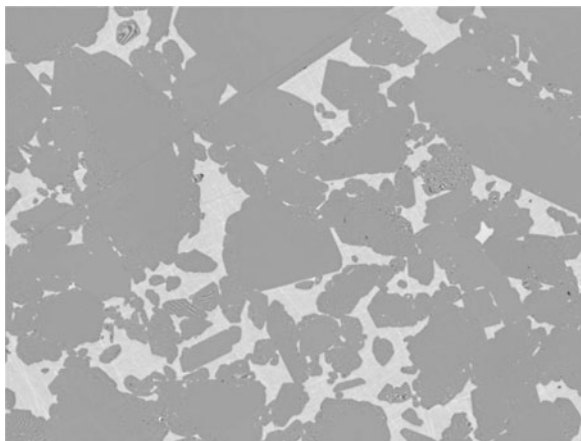
RBSiC for mechanical seals is available in two varieties: fine grained RBSiC and bimodal RBSiC. Fine grained RBSiC is manufactured from a relatively uniform grain SiC (commonly averaging between 8 and 10 μm) SiC material. This results in a relatively uniform distribution of SiC and Si in the final material (Figs. 11 and 12). Bimodal RBSiC (Fig. 13) starts with two different grain sizes. The different grain sizes are selected to improve packing density of the material and minimize secondary bonding and free silicon. Both materials are commonly available and both have proven to provide acceptable performance as a seal face material.

Self-Sintered Silicon Carbide

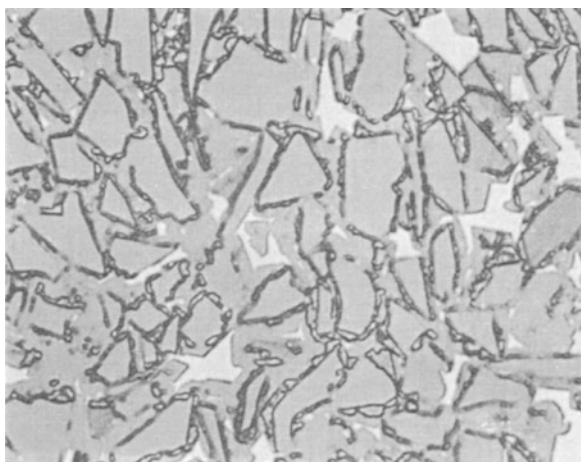
Researchers discovered that it was possible to sinter very fine grained SiC particles directly together using non-oxide sintering aids (such as C, B, or Al) in an inert



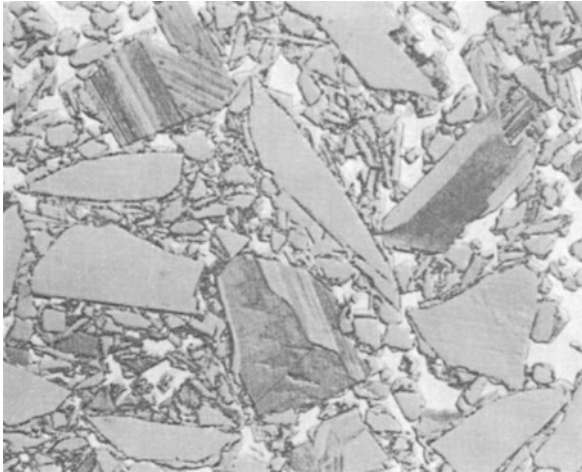
Materials for Mechanical Seals, Fig. 10 RBSiC tube after reaction bonding



Materials for Mechanical Seals, Fig. 11 RBSiC unetched, 100X



Materials for Mechanical Seals, Fig. 12 RBSiC etched, 100X

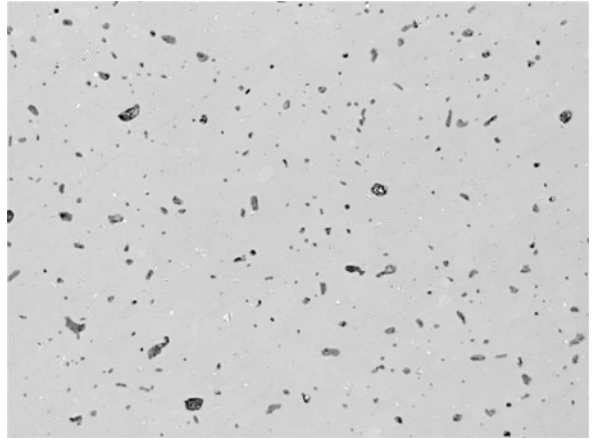


Materials for Mechanical Seals, Fig. 13 RBSiC bimodal, 100X

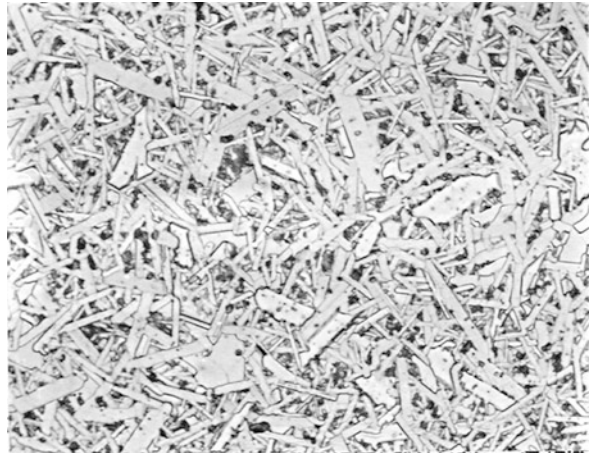
environment at temperatures over 2,000°C. The resulting material consists almost entirely of SiC (generally greater than 97% theoretical density) with a very small volume of unconnected voids (Figs. 14 and 15). Due to the lack of a secondary material (such as silicon), the direct sintered material is chemically resistant to almost any fluid and process condition likely to be seen in a centrifugal pump.

Some of the original research on self-sintered silicon carbides (SSSiC) was done with a beta phase SiC. Later research used alpha SiC. These new materials and manufacturing methods were patented, which greatly limited their availability and slowed their acceptance into industry. There was also a perception that SSSiC was more brittle and prone to chipping than RBSiC. Since there were no standardized tests for evaluating these properties, it remained as a perception with no conclusive data to prove one position or the other. The inert nature of the material, however, was required for many applications in the chemical manufacturing industry, where it became an informal industry standard. This material is now widely available and accepted in mechanical seals across all industries.

SSSiC is manufactured in a sintering process. In this process, SiC powder (along with binding agents) is pressed into a mold at high pressure. When it is removed from the mold in the green state, the SiC can be easily machined with diamond tooling. The mold may be in the shape of a tube or, if there is high volume of identical components being manufactured, into a mold that has the shape and features of the finished part. During the sintering operation, the SiC powder fuses with itself and becomes a solid piece of SiC. In this process, the part shrinks significantly (approximately 18% linearly – Fig. 16). Because this



Materials for Mechanical Seals, Fig. 14 SSSiC unetched, 100X



Materials for Mechanical Seals, Fig. 15 SSSiC etched, 400X



Materials for Mechanical Seals, Fig. 16 Shrinkage of SSSiC during sintering

shrinkage is predictable and repeatable, the manufacturer can machine or mold the part in the green state to a larger size and have it shrink predictably to the required final dimensions.

Self-sintered silicon carbide is often called sintered silicon carbide (SSiC), direct-sintered silicon carbide (DSSiC), or alpha sintered silicon carbide (α SiC). In the mechanical seal industry, these terms all indicate the same material and are used interchangeably. Of the two basic material types mentioned so far, SSSiC and RBSiC, testing has shown that RBSiC vs. carbon provides a better tribological pair from a wear standpoint.

Silicon Carbide with Controlled Features

The vast majority of SiC material is solid material free from any significant pores, voids, or imperfections. This results in seal faces that are lapped very flat and smooth. Under certain conditions, faces that are too flat or too smooth can prevent fluid from migrating across the seal faces. This can prevent proper lubrication of the sealing interface. Over the past several decades, SiC manufacturers have been creating SiC materials that have controlled features throughout the structure of the materials. When a seal face is lapped flat, these features create small anomalies on the surface that can provide enhanced seal performance.

The most common varieties of these materials involve free graphite particles dispersed throughout the SiC material. On a lapped surface, this results in small patches of graphite surrounded by a hard SiC surface. The graphite does not provide additional lubrication by adding graphite into the sliding interface. Rather, the soft graphite material creates small depressions in the surface that enhance lubrication with the process fluids. As an alternate approach, SiC material is available with small controlled voids throughout the material, which create small depressions on the lapped surfaces. It is even possible to create material with both free graphite and controlled porosity.

While these materials have proven to have some useful characteristics, there are also some limitations. The dispersion of porosity in the material is created during the mixing and pressing of the original raw materials. With good manufacturing techniques the free graphite and/or pores are usually evenly dispersed across the seal although their exact location and density is more a function of statistical probability than design. During operation, some portion of the free graphite present on the sealing surface is lost into the fluid film between the faces. While the resulting depressions contribute to the beneficial hydrodynamic effects, they are not constant and change

as more graphite is lost from the face. This can result in unexpected changes in seal performance over time. Finally, the presence of soft material or pores in the seal interface can lead to abrasive particles in the process fluid collecting in these locations. The resulting captured material can lead to aggressive face wear. Even with these limitations, these materials continue to be used in seal applications, although they are typically limited only to applications where additional fluid film formation is required.

Carbon Conversion Process – Siliconized Graphite

All of the SiC manufacturing methods described above begin with SiC powder as one of the raw ingredients. It is, however, possible to create SiC directly from carbon graphite in the final component (Fig. 17). When a solid carbon component is exposed to silicon monoxide vapors under controlled conditions, the carbon will react with the silicon vapors to form SiC. This reaction converts the outer layer of the material into SiC. This conversion however generally only penetrates a short distance into the carbon (typically 0.5–2.0 mm) and results in a SiC shell around the carbon core (Fig. 18). This material is porous by nature since this is necessary for the infusion of reactive Si gas. After the reaction is completed, remaining porosity can be impregnated in a manner similar to carbon seal faces.

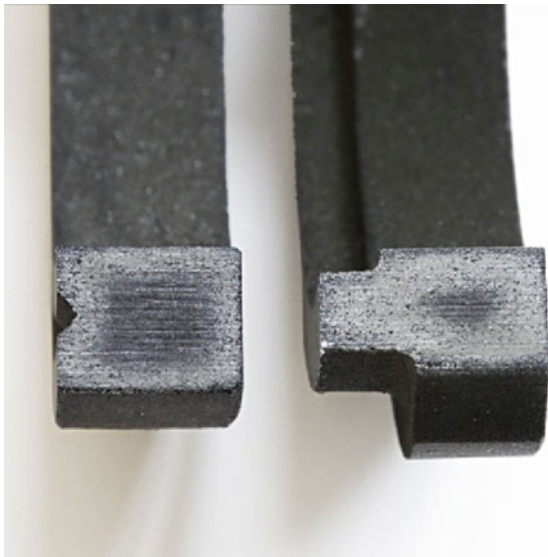
This is a relatively simple and inexpensive method of manufacturing SiC since almost all of the features of the component can be machined or pressed into the component while the part is made of carbon. This method of manufacturing, however, produces components with inferior mechanical properties (compared to other SiC options) since the final component has both SiC and carbon in the structure. It has, however, been used successfully for high volume components used in low duty services.

Silicon Nitride

Silicon nitride (SiN) was discovered in 1857 but did not find popularity until the 1950s. SiN is a chemical compound of silicon and nitrogen and is formed at temperatures between 1,000°C and 1,500°C depending on the particular manufacturing process. It has high strength over a wide range of temperature but with a thermal conductivity that is lower than SiC. Applications that use SiN in mechanical seals involve those where strength is the primary concern, such as in high speed compressor seals. In these applications where the hard face must



Materials for Mechanical Seals, Fig. 17 Small seal faces manufactured by the carbon conversion process

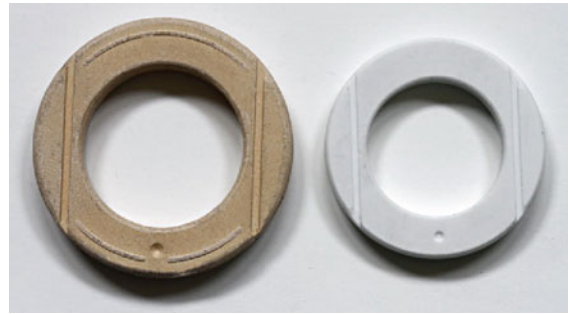


Materials for Mechanical Seals, Fig. 18 Silicon carbide shell around carbon core

rotate at high speeds, rotational stresses make the use of SiC improbable.

Tungsten Carbide

Tungsten carbide (WC) is a carbide ceramic that is used in many products requiring high hardness and toughness. WC shares many of the same difficulties in manufacturing shapes with other ceramics. WC is readily available in powder form but must be processed into a final shape. Tungsten carbides are most often manufactured as cemented carbides. As a cemented carbide there is no



Materials for Mechanical Seals, Fig. 19 Alumina oxide shrinkage during sintering

attempt to bond WC to itself. Rather a secondary metal is added to metallurgically bind or cement the WC particles together. This results in a material that has the combined properties of both the WC and the metal binder. This has been used to an advantage by providing greater toughness and impact strength than possible with WC alone. Although WC is classified as a brittle material (since it has almost no ductility), it is used in applications where the material is exposed to impacts or tensile loads. One of the primary weaknesses of cemented WC in mechanical seal applications is its high density.

Tungsten carbide powder is created by a reaction between tungsten metal and carbon in a hydrogen environment at approximately 1,450–1,500°C. The powder is then mixed with the metal binder (e.g., Co or Ni) and a small amount of paraffin. This mixture is placed into a mold and compressed at high pressure. The mold is generally in a shape close to the required shape of the seal face so machining and grinding is minimized. A pre-sintering heating cycle may be done followed by additional shaping. The part is then fired in a furnace at temperatures of 1,350–1,550°C. Any additional shaping must be done by grinding with diamond tooling.

Cobalt bound WC has proven to be one of the most successful carbides. These were introduced in the early 1930s for cutting tools in the metal working industry. As the mechanical seal developed over the years, Co-WC was the standard ceramic material. Cobalt though did not exhibit the range of chemical compatibilities required by industry and was gradually replaced by nickel bound WC. Current grades range from 6% to 10% free Nickel. Ni-WC is still widely used as a seal face material, especially where its high strength and high toughness properties are beneficial. It has good chemical compatibility but is generally limited by the free nickel. Ni-WC is often used in applications where two hard faces are required. Generally it is

Materials for Mechanical Seals, Table 1 Typical physical properties of seal face materials

	Resin Carbon	Antimony Carbon	RBSiC	SSSiC	Ni-WC (6% Ni)	Alumina (99.50%)	Silicon Nitride
Expansion coefficient $10^{-6}/^{\circ}\text{C}$	4.9–5.6	6.3	4.5	4.0–4.5	5.2	8.2	3.4
Thermal conductivity W/m K	9–14	14	110–151	80–153	82	35.6	20
Young's modulus GPa	21–24	26	360–393	400–413	607	372	290
Flexural strength MPa	76–79	96	345–462	380–550	–	379	500
Hardness	90–95 Scleroscope	95 Scleroscope	24.5 Gpa	27.5 Gpa	90–91 Ra	14.1 Gpa	13.7 Gpa
Density g/cc	1.8	2.2	3.1	3.1	14.9	3.9	3.2

mated against a SiC face. However, WC faces do not tolerate dry-running conditions or high face loading since this can result heat checking or cracking on the seal face.

Alumina Oxide

Oxides of aluminum range from oxidation on aluminum metals to rubies and sapphires. They also constitute the most widely used oxide ceramic in the form of alumina oxide (Al_2O_3), commonly referred to as alumina. Alumina is a white or ivory colored ceramic that can be produced into shapes by direct sintering. This results in a dense, homogeneous material. It shares many of the favorable attributes of other ceramics such as high hardness, high strength, and high stiffness. It is also has excellent dielectric properties. It is commonly used for electrical insulators, wear resistant components, grinding media, and high temperature components.

Alumina is available in a range of purities, typically from around 94% to over 99%. In the high purities, alumina has excellent chemical resistance to most process fluids except for some strong acids. Alumina does have one weakness that has restricted its use in seals; it can fracture easily if thermally shocked. Thermal shock may result from rapid heating in the process fluid or from dry-running operation. Even with this concern, it continues to be used where thermal shock is not considered an issue.

Alumina faces are manufactured in a sintering process similar to self-sintered SiC. Because of this, Alumina faces manufactured in higher volumes can be pressed into a mold into a near net shape. Detailed features can easily be added in the mold, which can reduce manufacturing cost of the faces. During sintering, there is significant shrinkage (Fig. 19) in the component but it is very predictable and the manufacturer can hold very tight tolerances on the fired part.

Comparison of Physical Properties

By reviewing the material available for seal faces, it is clear that each has its own set of strengths and weaknesses. It is not possible to declare one material superior in all applications. The demands of the process conditions, seal component design, tribological properties, and economic considerations all play a part in this evaluation. It is, however, possible to compare physical data, which can illustrate differences between the materials.

Data on physical properties and chemical compatibility shown below (Table 1) have been collected from a number of different sources, including the material manufacturers' published data. In some cases data may not be directly comparable due to different assumptions, sample preparation, sample purity, manufacturing processes, or test techniques. In some cases, this can lead to a wide range of properties within the same material type. Although all of the data in this chapter is believed to be correct, neither the authors, nor any other entity associated with this chapter, makes any claim to its use or accuracy. Users of materials of any nature are urged to obtain specific information from the manufacturer about its applicability in a specific application.

Cross-References

- [Mechanical Seal Analysis](#)
- [Mechanical Seals](#)

References

- S. Chinowsky, Friction and wear of carbon-graphite materials, in *ASM Handbook, Friction, Lubrication, and Wear Technology*, vol. 18, ed. by P. Blau (1992)
- M.B. Huebner, Material selection for mechanical seals, in *Proceedings of the 21st International Pump User's Symposium*, Turbomachinery Laboratory, Texas A&M University, College Station, Texas (2005)
- A.J. Massaro, The 'Mating Pairs' concept for mechanical seals. *Lubr. Eng.* **92**, 436–446 (1988)

- R.R. Paxton, *Manufactured Carbon: A Self-Lubricating Material for Mechanical Devices* (CRC Press, Boca Raton, 1979)
- Pumps – Shaft Sealing Systems for Centrifugal and Rotary Pumps*, API 682, 3rd edn. (American Petroleum Institute, Washington, DC, 2004)
- S.J. Schneider Jr. (ed.), *Engineered Materials Handbook*, vol. 4. Ceramics and Glasses (ASM International, Metals Park, Ohio, 1991)
- L.J. Thorwart, Select optimum face materials to ensure seal performance. *Chem. Eng. Prog.* **92**(4), 80–83 (1996)

Mathematical Foundation of Fluid Lubrication Theory

CODA H. PAN

Global Technology, Millbury, MA, USA

Synonyms

Hydrodynamic lubrication; Theory of fluid lubrication; Thin-film lubrication flow

Definition

Fluid lubrication theory expounds the mechanism of separating bodies in contact with a fluid film according to the physical principle of its motion through a thin gap of separation.

Scientific Fundamentals

Mechanism

If two bodies are pressed together, there can be friction to resist their relative sliding. Fluid lubrication fills the separating gap with a fluid film that develops an elevated pressure to resist the closing force. Generation of the pressure in the fluid lubricant, its relationship with the profile of film thickness, the associated shear stress at the surfaces, and the role of lubricant rheology are issues in lubrication theory.

While free or low-friction sliding between the surfaces is made possible, the precise geometry of one surface serves to constrain the location of the opposing body in rather precise terms. The overall dimension of the area, in which pressure is elevated to maintain the separation, is typically 1,000 or more times the separation gap; the reciprocal of this disproportionally large number is called the clearance ratio, a prominent parameter in fluid lubrication. Depending on the kinematic function of these bodies in contact, the surface geometry may not be a flat plane; examples of curved surfaces include the cylindrical surface between a shaft and the mating bushing, mated

conical frustums defining an axis while carrying a thrust load, the free rotating ball joint, and elements that experience considerable elastic deformation due to hydrodynamic pressure.

Conservation Laws, Rheology, and Newtonian Law of Viscosity

The fluid lubricant obeys the physical laws of continuum mechanics comprising conservation of mass, momentum, and energy. Rheology deals with the stress–strain relationship of the lubricant in motion. It is an input to the physical conservation laws that affects the final governing equations used in engineering computation of fluid lubrication systems. Many different aspects of rheology are encountered in fluid lubrication; some take on dominating roles and become identified with specialty topics of considerable importance. Rheology-dominated fluid lubrication topics are treated elsewhere in the Encyclopedia (► [Lubrication with a Non-Newtonian Fluid](#); ► [Lubrication with a Grease](#)). The ensemble of the three conservation laws of fluid mechanics is attributed to the writings of Navier (1823) and Stokes (1845). Reynolds (1886) first published the thin-film approximation of these equations toward an understanding of the origin of elevated pressure in a lubricating film, which is commonly identified with the theory of fluid-film lubrication. Generalization of lubrication theory to consider gas as a compressible lubricant was due to Harrison (1913), who recognized its isothermal attribute. Derivation of the governing equation for a lubricating fluid film, which emphasizes the continuity condition of the lubricating fluid, will be thoroughly examined.

Thin Film Continuum Physics

The lubricant enclosed between the separating surfaces adheres to both walls. In the presence of wall motion, it flows according to the physical laws of continuum mechanics in a very special way. The mathematical foundation of fluid lubrication lays out a rigorous examination of the consequence of a very small ratio of the film gap to the global dimension of the bearing.

Conservation Laws of Continuum Physics

The conservation laws of continuum physics present mathematical statements in the form of partial differential equations that universally govern the mechanics of flow.

Conservation of Mass The flow velocity \vec{v} and fluid mass density ρ are required to satisfy the continuity

condition dictating that the space is filled by the fluid in motion

$$\frac{\partial \rho}{\partial t} + \bar{\nabla} \cdot (\rho \bar{V}) = 0 \quad (1)$$

Conservation of Momentum Newton's law for dynamic equilibrium of flowing matter concerns the actions of pressure, body force, and internal stress in balance with flow momentum:

$$\begin{aligned} -\bar{\nabla} p + \bar{\nabla} \cdot \hat{\tau} + \rho \bar{G} &= \rho \left[\frac{\partial \bar{V}}{\partial t} + (\bar{\nabla} \times \bar{V}) \times \bar{V} \right. \\ &\quad \left. + \frac{1}{2} \bar{\nabla}(\bar{V} \cdot \bar{V}) \right] \\ &= \rho \left(\frac{\partial}{\partial t} + \bar{V} \cdot \bar{\nabla} \right) \bar{V} \end{aligned} \quad (2)$$

\bar{G} is the body force density that would be the earth gravity or its combination with platform acceleration, p is the fluid pressure, and $\hat{\tau}$ is the stress tensor. The right-hand side presents fluid inertia in two alternative ways; the top version brings out curvature effects that would be present for curved bearing surfaces, the lower one shows the effect of convective momentum gradient.

The stress tensor is defined as the part of the internal force that causes deformation; in particular, the mean normal force density is judged to be non-deforming and is excluded from the stress tensor. The stress tensor $\hat{\tau}$ is mathematically associated with two directions; the first one is the direction of the surface force, the second one is the normal direction of an area element, and the coefficient associated with the direction-pair is the surface force intensity per unit area. It is commonly represented by its Cartesian matrix elements:

$$[\tau] = \begin{bmatrix} \tau_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \tau_{yy} & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \tau_{zz} \end{bmatrix} \quad (3)$$

The sum of the three diagonal terms of the stress matrix is its first invariant. It must vanish because the non-deforming component has been excluded:

$$\tau_{xx} + \tau_{yy} + \tau_{zz} = 0 \quad (4)$$

As such, pressure and stress tensor appear as separate terms in the left-hand side of (2).

The stress tensor must have cross-diagonal symmetry to maintain rotational equilibrium at an infinitesimal scale:

$$\begin{aligned} \tau_{yx} &= \tau_{xy} \\ \tau_{zx} &= \tau_{xz} \\ \tau_{zy} &= \tau_{yz} \end{aligned} \quad (5)$$

Equations 3–5 together allow the state of stress to have five independent stress coefficients.

Equation 3 is a coordinate-dependent description of the stress tensor written in Cartesian coordinates with a specific orientation of its base vectors. Given a set of these five stress coefficients, it is always possible to find a transformation to reduce the stress matrix to a diagonal form. There are three orientations such that there is only normal stress on each face that is normal to the base vector. The normal stresses along the principle directions are called principle stresses; they would have a null average so that the pressure is not affected by the transformation:

$$\begin{aligned} [\tau'] &= \begin{bmatrix} \tau'_I & 0 & 0 \\ 0 & \tau'_{II} & 0 \\ 0 & 0 & \tau'_{III} \end{bmatrix} \\ \tau'_I + \tau'_{II} + \tau'_{III} &= 0 \end{aligned} \quad (6)$$

Adjunct to the principal directions, there are orientations of the base vectors that the base faces do not have normal stress; these may be called the shear directions:

$$[\tau''] = \begin{bmatrix} 0 & \tau''_I & \tau''_{II} \\ \tau''_I & 0 & \tau''_{III} \\ \tau''_{II} & \tau''_{III} & 0 \end{bmatrix} \quad (7)$$

Conservation of Energy Energy conservation in a flowing fluid equates the increase of enthalpy to heat generation by the deformation work of internal stresses:

$$\rho \left(\frac{\partial}{\partial t} + \bar{V} \cdot \right) h - \bar{\nabla} \cdot (\kappa \bar{\nabla} T) = \hat{\tau} \cdot \hat{V} - p \bar{\nabla} \cdot \bar{V} \quad (8)$$

\hat{V} represents the rate of deformation, it is a tensor with six independent elements:

$$[\hat{V}] = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) & \frac{1}{2} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \\ \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) & \frac{\partial v}{\partial y} & \frac{1}{2} \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) \\ \frac{1}{2} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) & \frac{1}{2} \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) & \frac{\partial w}{\partial z} \end{bmatrix} \quad (9)$$

$\hat{\tau} \cdot \hat{V}$ is the “deformation dissipation” that accounts for heat generated in the fluid due to deformation; in Cartesian notation, it is

$$\begin{aligned} \hat{\tau} \cdot \hat{V} &= \tau_{xx} \frac{\partial u}{\partial x} + \tau_{yy} \frac{\partial v}{\partial y} + \tau_{zz} \frac{\partial w}{\partial z} + \frac{1}{2} (\tau_{xy} + \tau_{yx}) \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \\ &\quad + \frac{1}{2} (\tau_{yz} + \tau_{zy}) \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) + \frac{1}{2} (\tau_{xz} + \tau_{zx}) \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \end{aligned} \quad (10)$$

Three other velocity gradient elements are the component of the vorticity vector that represent rotation of the fluid element and may render interpretation of the right-hand side of (2) but play no role in heat generation; they are

$$\bar{\nabla} \times \bar{V} = \bar{i} \left(\frac{\partial w}{\partial y} - \frac{\partial y}{\partial z} \right) + \bar{j} \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right) + \bar{k} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \quad (11)$$

Thin Film Scaling of Conservation Laws of Continuum Physics

Because the clearance ratio is numerically very small, mass conservation plays the most dominant role.

Incompressible Fluid Film Considering first the incompressible fluid, rewriting in Cartesian coordinates with the lower bearing surface coincident with the (x, y) plane, continuity requires

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \quad (12)$$

Let R and C be used respectively to scale “in-plane” and normal distances and let V_s be a representative velocity that can be estimated from known surface sliding or the mean flow that would be discharged from the periphery of a pressurized fluid film. Then, the first two terms in (12) can be estimated as

$$\left| \frac{\partial u}{\partial x}, \frac{\partial v}{\partial y} \right| \approx O \left\{ \frac{V_s}{R} \right\} \quad (13)$$

Equation 12 can be used to estimate the magnitude of the velocity component aimed along \bar{k} since it results from balancing the “in-plane” velocity gradients:

$$w \approx O \{ V_s C / R \} \quad (14)$$

A small clearance ratio has far reaching implications on the stress tensor. Consider first its appearance under the divergence operator in the left-hand side of (2). Again, using Cartesian coordinates, this is

$$\begin{aligned} \bar{\nabla} \cdot \bar{\tau} = & \bar{i} (\partial \tau_{xx} / \partial x + \partial \tau_{xy} / \partial y + \partial \tau_{xz} / \partial z) \\ & + \bar{j} (\partial \tau_{yx} / \partial x + \partial \tau_{yy} / \partial y + \partial \tau_{yz} / \partial z) \\ & + \bar{k} (\partial \tau_{zx} / \partial x + \partial \tau_{zy} / \partial y + \partial \tau_{zz} / \partial z) \end{aligned} \quad (15)$$

Order of magnitude estimation for the partial-spatial derivatives can continue as

$$\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \approx O \left\{ \frac{1}{R} \right\}, \text{ and } \frac{\partial}{\partial z} \approx O \left\{ \frac{1}{C} \right\}$$

Equation 15 immediately simplifies to

$$\bar{\nabla} \cdot \bar{\tau} \approx \bar{i} \partial \tau_{xz} / \partial z + \bar{j} \partial \tau_{yz} / \partial z + \bar{k} \partial \tau_{zz} / \partial z \quad (16)$$

Then Equation 2 is reduced to

$$\bar{\nabla} p = \bar{i} \frac{\partial \tau_{xz}}{\partial z} + \bar{j} \frac{\partial \tau_{yz}}{\partial z} + \rho \bar{G} - \rho \left(\frac{\partial \bar{V}}{\partial t} + \bar{V} \cdot \bar{\nabla} \right) \bar{V} \quad (17)$$

Since the above formula does not depend on any specific constitutive model, it can be used to deal with issues like viscosity-to-plasticity transition. One can introduce a scale for the shear stresses, say T , without the restrictive conditions of any rheological model, then add a generic gravity constant g and an unspecified time scale Δt to estimate the relative importance of the right-hand side of the above equation if the pressure in the fluid film is computed by integrating from an ambient edge (Table 1).

(R/C) can be interpreted as the lift-to-drag ratio of the lubricating fluid film. If a target of load per unit area is available, say Ψ , then an estimation of the operating clearance ratio is T/Ψ . A large target of load capacity calls for a small clearance ratio. Designer's choice would be a compromise with manufacturing cost consideration; typically, the clearance ratio is in the range of 10^{-3} – 10^{-4} . The bearing surface would be practically aligned with the shear planes that present the stresses as described by (7); the important stress components $(\tau_{xz}, \tau_{yz}, \tau_{zz})$ are surface forces on the z -face and can be denoted with single subscripts as surface forces $(\bar{i} \tau_x, \bar{j} \tau_y, \bar{k} \tau_z)$.

If one is willing to neglect $\rho \bar{G} - \rho (\partial \bar{V} / \partial t + \bar{V} \cdot \bar{\nabla}) \bar{V}$ (17) can be integrated in the direction of \bar{k} to any point of the lubricant film to obtain

$$\bar{\tau}(z \leq h) = \bar{\tau}(0) + \int_0^z (\bar{\nabla} p) dz \quad (18)$$

The Newtonian law of viscosity (► [Lubrication with a Non-Newtonian Fluid](#); ► [Lubrication with a Grease](#)) postulates a homogeneous linear relationship between stress and rate of deformation tensors. If there is no directional property in the fluid property, then Newton's viscosity law can be expressed in two formulas; one for the shear stress, the other for the normal stress

Mathematical Foundation of Fluid Lubrication Theory,**Table 1** Contribution to fluid film pressure

Term	Order of magnitude contribution to $p - p_a$
$\partial\tau_{xz}/\partial z, \partial\tau_{yz}/\partial z$	$T(R/C)$
$\rho \bar{G}$	$\rho g R$
$\rho \partial \bar{V}/\partial t$	$\rho(V_s/\Delta t)$
$\rho \bar{V} \cdot \bar{\nabla}$	ρV_s^2

$$\begin{aligned}
 (\tau_{xy}, \tau_{yz}, \tau_{zx}) &= \mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}, \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) \\
 (\tau_{xx}, \tau_{yy}, \tau_{zz}) &= 2\mu \left(\frac{\partial u}{\partial x}, \frac{\partial v}{\partial y}, \frac{\partial w}{\partial z} \right)
 \end{aligned} \quad (19)$$

Neglecting terms of $o\{C/R\}$ then substitute into (18); it is found

$$\begin{aligned}
 \mu \frac{\partial u}{\partial z} &= \tau_{x,0}(0) + \frac{\partial p}{\partial x} z \\
 \mu \frac{\partial v}{\partial z} &= \tau_{y,0}(0) + \frac{\partial p}{\partial x} z \\
 2\mu \frac{\partial w}{\partial z} &= \tau_{z,0}(0) + \int_0^z \frac{\partial p}{\partial z} dz
 \end{aligned} \quad (20)$$

In view of the small clearance condition, the last line indicates $p - p_0$ is $o\{C/R\}$ and can be neglected; hence $\bar{\nabla} p$ can be treated as uniform across the film thickness and (18) can be reduced to the in-plane problem with

$$\bar{\nabla} \approx \left(\bar{i} \frac{\partial}{\partial x} + \bar{j} \frac{\partial}{\partial y} \right) + o\{C/R\}$$

Assuming a stationary lower surface and non-slip adhesion of the fluid to both enclosing surfaces and integrating across the film thickness, it is obtained,

$$\begin{aligned}
 \text{Velocity Gradient: } \frac{\partial \bar{V}}{\partial z} &= \frac{\bar{V}_h}{h} + \mu^{-1} \left(z - \frac{1}{2}h \right) \bar{\nabla} p \\
 \text{Velocity Profile: } \bar{V} &= \frac{z}{h} \bar{V}_h + (2\mu)^{-1} z(z-h) \bar{\nabla} p \\
 \text{Film Flux: } \bar{\Phi} &= \frac{1}{2} \bar{V}_h h - (12\mu)^{-1} h^3 \bar{\nabla} p
 \end{aligned} \quad (21)$$

To remove the restrictions of Newton's viscosity law as contained in (19), one only needs to allow the viscosity coefficient be dependent on z . Then, a generalization of (20) can be readily written out as

$$\begin{aligned}
 \text{Velocity Gradient: } \frac{\partial \bar{V}}{\partial z} &= \frac{\bar{V}_h}{\mu \int_0^h \mu^{-1} dz} \\
 &+ \left(z - \frac{\int_0^h \mu^{-1} z dz}{\int_0^h \mu^{-1} dz} \right) \frac{\bar{\nabla} p}{\mu}
 \end{aligned}$$

$$\begin{aligned}
 \text{Velocity Profile: } \bar{V}(z) &= \frac{\int_0^h \mu^{-1} dz}{\int_0^h \mu^{-1} dz} \bar{V}_h \\
 &- \bar{\nabla} p \left[\frac{\int_0^h \mu^{-1} z dz}{\int_0^h \mu^{-1} dz} \int_0^z \mu^{-1} dz \right. \\
 &\quad \left. - \int_0^z \mu^{-1} z dz \right]
 \end{aligned}$$

$$\begin{aligned}
 \text{Film Flux: } \bar{\Phi} &= \frac{\int_0^h \int_0^h \mu^{-1} dz^2}{\int_0^h \mu^{-1} dz} \bar{V}_h \\
 &- \bar{\nabla} p \left[\frac{\int_0^h \mu^{-1} z dz}{\int_0^h \mu^{-1} dz} \int_0^h \int_0^h \mu^{-1} dz^2 \right. \\
 &\quad \left. - \int_0^h \int_0^h \mu^{-1} z dz^2 \right]
 \end{aligned} \quad (22)$$

Although stationary surfaces were used to derive the expression for $\bar{\Phi}$, Equation (12) can be integrated between moving bearing surfaces to derive the governing equation for an incompressible lubricating film:

$$\bar{\nabla} \cdot \bar{\Phi} + \frac{\partial h}{\partial t} = 0 \quad (23)$$

Fluid viscosity is sensitive to temperature rise due to viscous heating within the fluid. When steady state operation is reached, temperatures of fluid film surfaces would be nearly uniform. If prior experience of the steady state temperature is available, an iso-viscous computation of (23) is usually adequate. If a more precise knowledge of fluid film temperature is desired, (8) would have to be calculated; the small clearance ratio condition makes possible much simplification of its right-hand side:

$$\begin{aligned}
 \rho \left(\frac{\partial}{\partial t} + \bar{V} \cdot \bar{\nabla} \right) h - \bar{\nabla} \cdot (\kappa \bar{\nabla} T) &= \mu \left[\left(\frac{\partial u}{\partial z} \right)^2 + \left(\frac{\partial v}{\partial z} \right)^2 \right] \\
 &- p(\bar{\nabla} \cdot \bar{V})
 \end{aligned} \quad (24)$$

Computation of this equation involves simultaneous treatment of convection of film flow and conduction to the surfaces; it is a typical parabolic-type partial differential equation that involves tracing the fluid motion as the

temperature across the film thickness is being determined. Upon knowing the temperature profiles, viscosity-temperature data of the fluid would be used to make possible computation of various moment integrals of the reciprocal of viscosity in (22).

Gas Film Lubrication Gas film lubrication came to prominence in the middle of the twentieth century when there was a need to install machinery in environments that could not tolerate contamination by lubricant vapors or vice versa.

The perfect gas law stipulates a simple relationship among the three thermodynamic state variables:

$$p = \rho \mathcal{R}T \quad (25)$$

Dependence of density on temperature (24), needs to be examined in terms of thermal transport properties of the gases of interest. The temperature of the surfaces can be treated as an unspecified level, say T . Heating, represented by the right-hand side, sets the level of temperature rise in the fluid film; the latter as a ratio to the reference level can be estimated as

$$\frac{\Delta T}{T} = \frac{\mu(V_s/C)^2}{(\kappa/C^2)T} = \frac{\mu C_p}{\kappa} (\gamma - 1) M^2$$

$\mu C_p/\kappa$ is the Prandtl number that characterizes convection heat transfer; it is $O\{1\}$ for common gases. M , the Mach number, is the ratio of flow or flight speed to the speed of sound; the latter is in the order of 10^2 m/s. V_s is typically less than 10 m/s, therefore, temperature rise due to local heating can be safely neglected. The other two terms account for transient temperature rise and convection heat transfer; their relative importance can be compared to the conduction term

$$\frac{|\rho \partial h / \partial t|}{|\bar{\nabla} \cdot (\kappa \bar{\nabla} T)|} \approx \frac{\rho V_s / \Delta t}{\mu V_s / C^2} \frac{\mu C_p}{\kappa}$$

$$\frac{|\rho \bar{V} \cdot \bar{\nabla} h|}{|\bar{\nabla} \cdot (\kappa \bar{\nabla} T)|} \approx \frac{\rho V_s^2 / R}{\mu V_s / C^2} \frac{\mu C_p}{\kappa}$$

These are both seen to be negligible due to the small clearance condition and that $\mu C_p/\kappa$ is $O\{1\}$. In conclusion, an isothermal state prevails for the gaseous thin film; and by virtue of the perfect gas law, density can be replaced by pressure in both (1) and (2), and Newton's viscosity law can be safely accepted; otherwise the formulas derived for the incompressible fluid remain valid:

$$\begin{aligned} \text{Velocity Gradient: } \frac{\partial \bar{V}}{\partial z} &= \frac{\bar{V}_h}{h} + \left(z - \frac{1}{2}h\right) \frac{\bar{\nabla} p}{\mu} \\ \text{Velocity Profile: } \bar{V}(z) &= \frac{\bar{V}_h z}{h} + \frac{\frac{1}{2}z(z-h) \bar{\nabla} p}{\mu} \\ \text{Volume Flux: } \bar{\Phi} &= \frac{1}{2} \bar{V}_h h - \frac{h^3 \bar{\nabla} p}{12\mu} \\ \text{Pressure Flux: } p \bar{\Phi} &= \frac{1}{2} p \bar{V}_h h - \frac{h^3 p \bar{\nabla} p}{12\mu} \\ \text{Continuity: } \bar{\nabla} \cdot (p \bar{\Phi}) &+ \frac{\partial (p h)}{\partial t} = 0 \end{aligned} \quad (26)$$

Non-dimensional Formulation of Flux Laws

Being a derivative of fluid mechanics, it is natural to embark on non-dimensional formulation of fluid film lubrication analysis. Scaling common to both incompressible and gas films involves $(C, R; V_s)$, respectively, for $(h, \bar{\nabla}; C^{-1} \bar{\Phi})$ to define

$$\begin{aligned} H &= h/C \\ \text{grad} &= R \bar{\nabla} \\ \bar{\Psi} &= \bar{\Phi} / (V_s C) \end{aligned} \quad (27)$$

For velocity scaling, there is a distinction between an externally pressurized bearing and one that relies on surface sliding to cause film pressurization:

$$\begin{aligned} \text{Pressurized Liquid Bearing: } V_s &= \frac{Q}{\oint C d\ell} \\ &\quad \text{Discharge periphery of bearing} \\ \text{Sliding Liquid Bearing: } V_s &= \frac{1}{2} V_h \end{aligned} \quad (28)$$

The divergence operator for the continuity equation takes on different forms that are dependent on the natural coordinate system for the particular bearing geometry.

Shown in Fig. 2 are cross-pattern pressure profiles and induced transverse pressure gradient at various through flows computed at 10^5 rpm; moderate compressibility effects are revealed by curvature in the profiles. Figure 3 illustrates pressurization lines up to 10^6 rpm; In the following, the normalized forms for two most commonly used bearings are written out:

$$\begin{aligned} \text{Cylindrical Journal Bearing: } \text{div } \bar{\Psi} &= \frac{\partial \Psi_\theta}{\partial \theta} + \frac{\partial \Psi_\gamma}{\partial \zeta} \\ \text{Circular Thrust Bearing: } \text{div } \bar{\Psi} &= \frac{\partial \Psi_\theta}{\partial \theta} + \frac{R}{r} \frac{\partial (r \Psi_r)}{\partial r} \end{aligned} \quad (29)$$

Incompressible Fluid Film For an incompressible fluid, absolute pressure has no significance; film pressure elevation from the ambient condition would be used. The non-dimensional flux laws are

$$\begin{aligned} \text{Sliding Liquid Bearing: } P &= \frac{C^2(p - p_a)}{6\mu V_h R} \\ \bar{\Psi} &= \bar{i}_s \eta_C H - \eta_p H^3 \text{ grad } P \\ \text{Pressurized Liquid Bearing: } P &= \frac{C^2(p - p_a)}{12\mu V_s R} \\ \bar{\Psi} &= -\eta_p H^3 \text{ grad } P \\ \text{Non-Newtonian Factors: } \eta_C &= \frac{2 \int_0^h \int_0^h \mu^{-1} dz^2}{h \int_0^h \mu^{-1} dz} \\ \eta_p &= \frac{12\mu}{h^3} \left(\frac{\int_0^h \mu^{-1} z dz}{\int_0^h \mu^{-1} dz} \int_0^h \int_0^h \mu^{-1} dz^2 - \int_0^h \int_0^h \mu^{-1} z dz^2 \right) \end{aligned} \quad (30)$$

By scaling volume flux, its divergence operation and the film thickness as indicated above, the natural scale of time is found to be (C/V_s) . Equation 23 is made non-dimensional as

$$\text{div } \bar{\Psi} + 2 \frac{\partial H}{\partial \tau} = 0 \quad (31)$$

Gaseous Film The isothermal gas film has the same volume flux law but its absolute pressure is used to emulate density for mass flow continuity consideration. Film pressure has to be scaled by the ambient pressure, and the Couette-versus-Poiseuille flux feature is indicated by a new, dimensionless parameter known as the bearing number for sliding gas film bearings

$$\begin{aligned} P &= p/p_a, \quad A \equiv 6\mu V_h (p_a C^2)^{-1} R \\ \bar{\Xi} &= \bar{i}_s P H - \frac{H^3 P \text{ grad } P}{A} \end{aligned} \quad (32)$$

The continuity condition for a gas film is

$$\text{div } \left(\bar{\Xi} \right) + 2 \frac{\partial (PH)}{\partial \tau} = 0 \quad (33)$$

Appearance of (PH) as time-derivative in (33) gives the gas film a very special property; if its enclosing surfaces are forced to “squeeze” at a high frequency with an amplitude that is a non-negligible fraction of the mean separation, the ambient edge functions as an oscillating pump to maintain an elevated pressure in the interior of the gas film. This phenomenon can be utilized to maintain separation of the surfaces against a moderate load level.

For externally pressurized gas bearings, in the absence of sliding, the top line of (28) would be used for velocity scaling; the non-dimensional pressure flux of an externally pressurized gas bearing is

$$\bar{\Xi} = - \oint p_a C^3 d\ell / (12\mu R Q) \quad H^3 P \text{ grad } P \quad (34)$$

Discharge periphery of bearing

Classical and Some Useful Analytical Solutions

One-Dimensional Sliders Lord Rayleigh examined solutions of the slider without side-leakage (1918), which are special cases of (30) with various profiles of the slider gap $H(0 \leq x/\ell \leq 1)$:

$$\Phi = H - H^3 \frac{h_0^2}{6\mu VR} \frac{d(p - p_a)}{d(x/\ell)}$$

$$\text{Plane: } H = 1 + (h_i/h_0 - 1)(1 - x/\ell)$$

$$n\text{-th Power: } H = 1 + (h_i/h_0 - 1)[1 - (x/\ell)^n]$$

$$\text{Exponential: } H = \exp\{\ln[1 + (h_i/h_0 - 1)(1 - x/\ell)]\}$$

The generic slider solution is synonymous with the determination of a flux invariant Φ

$$\Phi = \int_0^1 H^{-2} d\xi / \int_0^1 H^{-3} d\xi \quad (35)$$

The specific profile functions make available closed-form formulas of the integrals that appear in (35). Lord Raleigh sought the optimum value of h_i/h_0 for a maximum load P and took the additional step of making use of the calculus of variations to establish the optimized Rayleigh step-slider. A summary of Lord Rayleigh’s curiosis excursion is listed in Table 2.

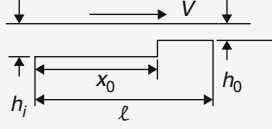
One-Dimensional Journal Bearing The celebrated Sommerfeld solution (1904) of a long, eccentric journal bearing is basically similar to (35). The half-angle transformation is utilized to integrate the reciprocal of H to yield the Sommerfeld angle ψ

$$\begin{aligned} \psi &= (1 - \varepsilon^2)^{\frac{1}{2}} \int_0^1 (1 + \varepsilon \cos \theta)^{-1} d\theta \\ &= 2 \tan^{-1} \left\{ \sqrt{(1 - \varepsilon)/(1 + \varepsilon)} \tan \frac{1}{2} \theta \right\} \end{aligned} \quad (36)$$

It in turn makes possible subsequent integration of H^{-2} and H^{-3} :

$$\begin{aligned} \int (1 + \varepsilon \cos \theta)^{-2} d\theta &= (1 - \varepsilon^2)^{-\frac{3}{2}} (\psi - \varepsilon \sin \psi) \\ \int (1 + \varepsilon \cos \theta)^{-3} d\theta &= (1 - \varepsilon^2)^{-\frac{5}{2}} \left[\left(1 + \frac{1}{2} \varepsilon^2\right) \psi \right. \\ &\quad \left. - 2\varepsilon \sin \psi + \frac{1}{4} \varepsilon^2 \sin 2\psi \right] \end{aligned} \quad (37)$$

Mathematical Foundation of Fluid Lubrication Theory,
Table 2 Optimized sliders

Profile	Optimum h_i/h_0	Achieved P
Plane	2.2	0.160
2nd power	2.3	0.163
Exponential	2.3	0.165
Rayleigh step	1.87	0.206
$x_0/\ell = 0.7123$		

By employing the above formulas with the periodic boundary condition on $(p - p_a)$, the flux invariant for the full circle solution was found as

$$\begin{aligned}\Phi_\theta &= \left(\int_{\theta_{\text{reference}}}^{\theta_{\text{reference}}+2\pi} (1 + \varepsilon \cos \theta)^{-3} d\theta \right) \\ &\quad \int_{\theta_{\text{reference}}}^{\theta_{\text{reference}}+2\pi} (1 + \varepsilon \cos \theta)^{-2} d\theta \quad (38) \\ &= (1 + \tfrac{1}{2}\varepsilon^2)^{-1} (1 - \varepsilon^2)\end{aligned}$$

The corresponding film pressure would result in a net force that is directed from the convergent side toward the divergent side, normal to the line of centers with the value

$$\frac{C^2 P}{12\pi\mu\omega R^3} = \frac{\varepsilon}{\sqrt{1 - \varepsilon^2}(2 + \varepsilon^2)} \quad (39)$$

The reciprocal of the left-hand side is known as the Sommerfeld number S .

An important property of the flux law for the journal bearing is that a pressure extremum coincides with where the flux invariant equals the Couette flux. In the full-circle Sommerfeld solution, there is a pressure peak in the converging part of the film profile and a pressure valley in the diverging part of the film profile; they are symmetrically located on either side of the minimum film thickness point, respectively in the second and third quadrant

$$\theta_{\text{pressure extremum}} = \cos^{-1} \left[-\tfrac{3}{2}\varepsilon / (1 + \tfrac{1}{2}\varepsilon^2) \right] \quad (40)$$

With an unspecified lower limit of integration $\theta_{\text{reference}}$ to define the flux invariant, the full-circle Sommerfeld solution permits designation of the pressure level of one chosen $\theta_{\text{reference}}$; this lack of uniqueness has caused concern over whether Sommerfeld's classical solution fits reality in the industrial application of fluid-film lubrication that sub-ambient pressure in a lubricating fluid film

cannot be sustained (► [Cavitation Formation and Modeling](#)). The following are alternative views to examine the Sommerfeld solution in this regard:

- Since Sommerfeld assigned the lower limit of the three integrals to $\theta = 0$, the film pressure there would be ambient and the divergent semi-circle would be sub-ambient; presence of such sub-ambient condition is believed to be contradicted by experience, the latter view is generally attributed to Gumbel (1921).
- The film pressure must terminate simultaneously with a null gradient at some location beyond the minimum film thickness point. This is based on a stability argument, which was independently put forth by Swift (1932) and Stieber (1933).
- The full-circle Sommerfeld solution can be made compatible with the Swift-Stieber condition by biasing the minimum film pressure above the ambient to realize the arc span

$$\begin{aligned}\cos^{-1} \left[-\tfrac{3}{2}\varepsilon / (1 + \tfrac{1}{2}\varepsilon^2) \right]_{\text{3rd quadrant}} - 2\pi &\leq \theta \\ &\leq \cos^{-1} \left[-\tfrac{3}{2}\varepsilon / (1 + \tfrac{1}{2}\varepsilon^2) \right]_{\text{3rd quadrant}}\end{aligned}$$

Generalization of the Sommerfeld solution to a partial-arc, which still satisfies the Swift-Stieber condition, would have both the flux invariant and the exit location bracketed with upper bounds according to (38) and (40):

$$\begin{aligned}0 &< \Phi_\theta < (1 + \tfrac{1}{2}\varepsilon^2)^{-1} (1 - \varepsilon^2) \\ \pi &< \theta_{\text{exit}} < \cos^{-1} \left[-\tfrac{3}{2}\varepsilon / (1 + \tfrac{1}{2}\varepsilon^2) \right]_{\text{3rd quadrant}}\end{aligned} \quad (41)$$

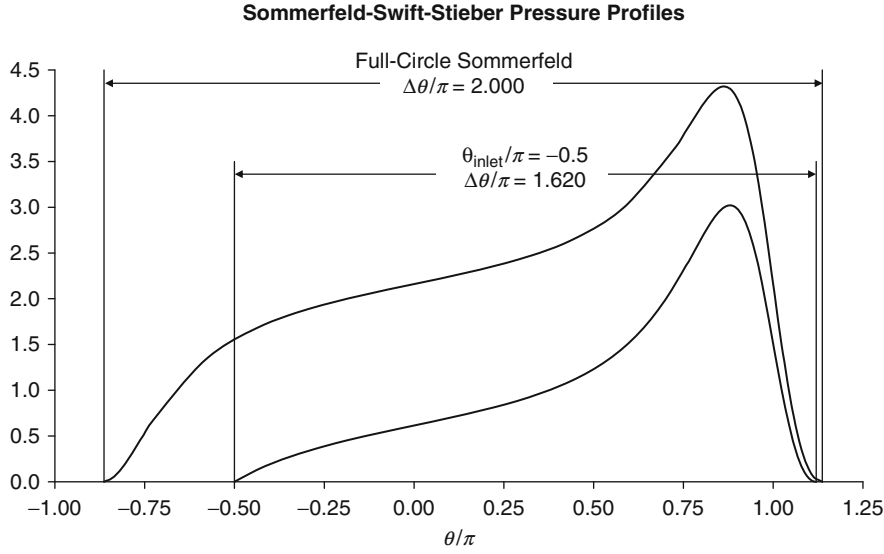
The film pressure profile for a partial-arc Sommerfeld solution with an assigned θ_{inlet} is calculated with an unspecified flux invariant Φ_θ :

$$\frac{C^2(p - p_a)}{6\mu\omega R^2} = \int_{\theta_{\text{inlet}}} H^{-2} d\theta - \Phi_\theta \int_{\theta_{\text{inlet}}} H^{-3} d\theta \quad (42)$$

The Swift-Stieber condition at θ_{exit} requires $\Phi_\theta = H_{\text{exit}}$; this can be satisfied by iterating $(\Phi_\theta, \theta_{\text{exit}})$ conjunctively according to (41). Examples of a partial-arc Sommerfeld-Swift-Stieber pressure profile and that of a biased full-circle pressure profile are shown in Fig. 1.

Short Journal Bearing Solutions Michell (1929) derived an analytical formula for the parabolic transverse pressure profile by ignoring the circumferential Poiseuille flow while balancing the eccentric Couette flux with the side leakage flux; thus

$$\begin{aligned}\frac{C^2(p - p_0)}{6\mu\omega R^2} &= -\tfrac{1}{2}H^{-3} \frac{\partial H}{\partial \theta} \zeta^2 \\ \frac{C^2(p_0 - p_a)}{6\mu\omega R^2} &= -\tfrac{1}{2}H^{-3} \frac{\partial H}{\partial \theta} \left(\frac{L}{D} \right)^2\end{aligned} \quad (43)$$



Mathematical Foundation of Fluid Lubrication Theory, Fig. 1 Full-circle and partial-arc Sommerfeld-Swift-Stieber pressure profiles

The above formula has been popularly used to model film pressure of short journal bearings with allowance for film rupture where the film thickness diverges $\pi \leq \theta \leq 2\pi$, this is known as the π -film; formulas of the projected film pressure integrals are derived with the aid of the Sommerfeld angle, there are both radial and tangential components:

$$W = \frac{(L/D)^2 \varepsilon \sqrt{(1 - \varepsilon^2)\pi^2 + 16\varepsilon^2}}{24\pi (1 - \varepsilon^2)^2} \quad (44)$$

$$\phi = \tan^{-1} \frac{\pi \sqrt{1 - \varepsilon^2}}{4 \varepsilon}$$

As an alternative to the π -film, (43) can be retained but film rupture can be defined by a rupture boundary as proposed by Pan and Ibrahim (1981) (► [Cavitation Formation and Modeling](#))

$$|\zeta_{\text{rupture}}| \leq (L/D) \quad (45)$$

Full-circle feeding of lubricant can be supplied to both ends of the bearing nominally at the ambient condition to treat the submerged short bearing, then the rupture condition would have to be at a sub-ambient level; or the lubricant can be supplied to one end slightly above the ambient level, then the rupture condition can be assumed to be at ambient. Both possibilities have been quite thoroughly treated. These results cannot be represented by closed form analytical formulas; hence, numerical results are tabulated.

Harrison's Solution of Air-Lubricated Sliders Exploring the significance of the bearing number, Harrison rearranged the isothermal flux law to make the product PH the primary unknown variable in the isothermal flux law

$$\Xi = \left[\bar{i}_s + \Lambda^{-1}(\text{grad } H)(PH) \right] PH - \Lambda^{-1} PH^2 \text{grad } PH \quad (46)$$

Turning attention to one-dimensional sliders with ambient boundary conditions, Harrison re-wrote (46) into the total differential form

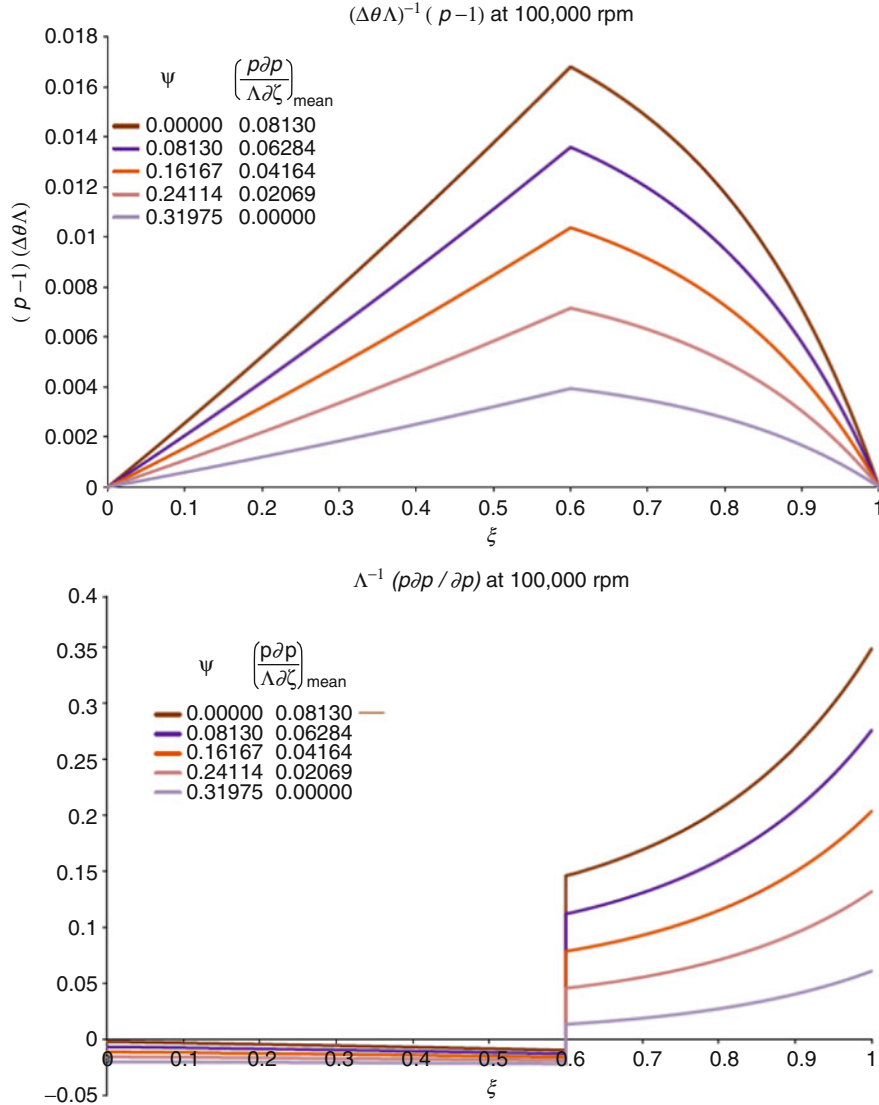
$$d\xi = \frac{PH^2 d(PH)}{(dH/d\xi)(PH)^2 + \Lambda(PH - \Xi)} \quad (47)$$

Analytical closed-form integrals of (47) were found for the step slider and the inclined plane slider.

In the case of step sliders, two transcendental formulas of $P(\xi, \Xi)$ are found for either side of the step with a single, undefined invariant Ξ :

$$H(0 \leq \xi \leq \xi_0 < 1) = H_1 > 1 : \Lambda \xi = H_i [(P - 1)H_i + \Xi \ln \frac{PH_i - \Xi}{H_i - \Xi}]$$

$$H(\xi_0 \leq \xi \leq 1) = 1 : \Lambda(\xi - 1) = P - 1 + \Xi \ln \frac{P - \Xi}{1 - \Xi} \quad (48)$$



Mathematical Foundation of Fluid Lubrication Theory, Fig. 2 Modest compressibility effect revealed in cross-pattern profiles in a miniature spiral groove thrust bearing

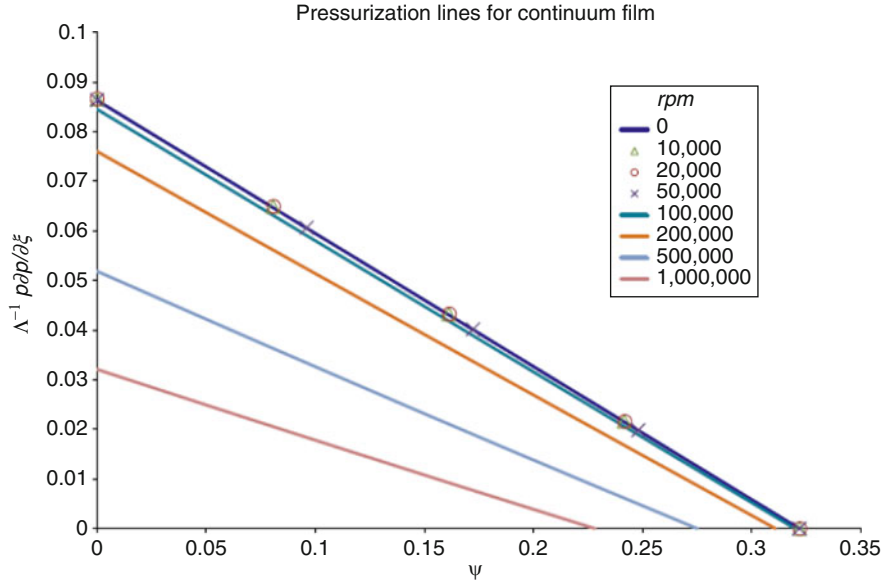
The values of (Ξ, P_0) are found by numerical iterations for given (Λ, ξ_0, H_i) between two transcendental equations resulting from assigning $\xi = \xi_0$ to (48):

$$\begin{aligned} \Lambda\xi_0 &= H_i \left[(P_0 - 1)H_i + \Xi \ln \frac{P_0 H_i - \Xi}{H_i} - \Xi \right] \\ \Lambda(\xi_0 - 1) &= P_0 - 1 + \Xi \ln \frac{P_0 - \Xi}{1 - \Xi} \end{aligned} \quad (49)$$

Afterwards, inverse pressure profiles can be calculated with (48). Such computations achieve very high resolution that is equivalent to use of dozens of computation mesh

intervals of a conventional finite-difference-method or finite-element-method. This was done to examine potential pressurization degradation of a high speed miniature spiral-groove thrust bearing, 2 mm OD up to 10^6 rpm, which features 18 grooves.

Shown in Fig. 2 are cross-pattern pressure profiles and induced transverse pressure gradient at various through flows computed at 10^5 rpm; moderate compressibility effects are revealed by curvature in the profiles. Figure 3 illustrates pressurization lines up to 10^6 rpm; severe degradation due to high bearing number is predicted at rotation rates beyond 10^5 rpm.



Mathematical Foundation of Fluid Lubrication Theory, Fig. 3 Degradation of spiral groove thrust bearing at very high speeds

Closed form integration of (47) results in a transcendental formula for determination of the flux invariant Ξ of the inclined plane slider; if $1 + 4(dH/d\xi)/\Lambda > 0$

$$\begin{aligned} & \ln \frac{1}{H_i} - \frac{1}{2} \ln \frac{(dH/d\xi) + \Lambda(1 - \Xi)}{(dH/d\xi)H_i^2 + \Lambda(H_i - \Xi)} \\ &= \frac{\frac{1}{2}\Lambda}{(dH/d\xi)\sqrt{1 + 4(dH/d\xi)/\Lambda}} \\ & \ln \frac{1 + \frac{1}{2}\Lambda/(dH/d\xi) + \sqrt{1 + 4(dH/d\xi)/\Lambda}H_i + \frac{1}{2}\Lambda/(dH/d\xi) - \sqrt{1 + 4(dH/d\xi)/\Lambda}}{1 + \frac{1}{2}\Lambda/(dH/d\xi) - \sqrt{1 + 4(dH/d\xi)/\Lambda}H_i + \frac{1}{2}\Lambda/(dH/d\xi) + \sqrt{1 + 4(dH/d\xi)/\Lambda}} \end{aligned} \quad (50)$$

Or, if $1 + 4(dH/d\xi)/\Lambda < 0$

$$\begin{aligned} & \ln \frac{1}{H_i} - \frac{1}{2} \ln \frac{(dH/d\xi) + \Lambda(1 - \Xi)}{(dH/d\xi)H_i^2 + \Lambda(H_i - \Xi)} \\ &= \frac{\frac{1}{2}\Lambda}{(dH/d\xi)\sqrt{1 + 4(dH/d\xi)/\Lambda}} \\ & \left[\tan^{-1} \frac{\sqrt{1 + 4(dH/d\xi)/\Lambda}}{1 + \frac{1}{2}\Lambda/(dH/d\xi)} - \tan^{-1} \frac{\sqrt{1 + 4(dH/d\xi)/\Lambda}}{H_i + \frac{1}{2}\Lambda/(dH/d\xi)} \right] \end{aligned} \quad (51)$$

The structure of (47) makes available an existence theorem that assures rapid iterative convergence for computation of Ξ , Jäger, C.H.T. Pan (1985). A full range bearing number calculation, $0 \leq \Lambda \leq \infty$, of the pressure profile that spans both signs of $1 + 4(dH/d\xi)/\Lambda$ for $H_i = 3.0$ is shown in Fig. 4.

A profile concatenation technique was used to adapt Harrison's plane slider integral to model

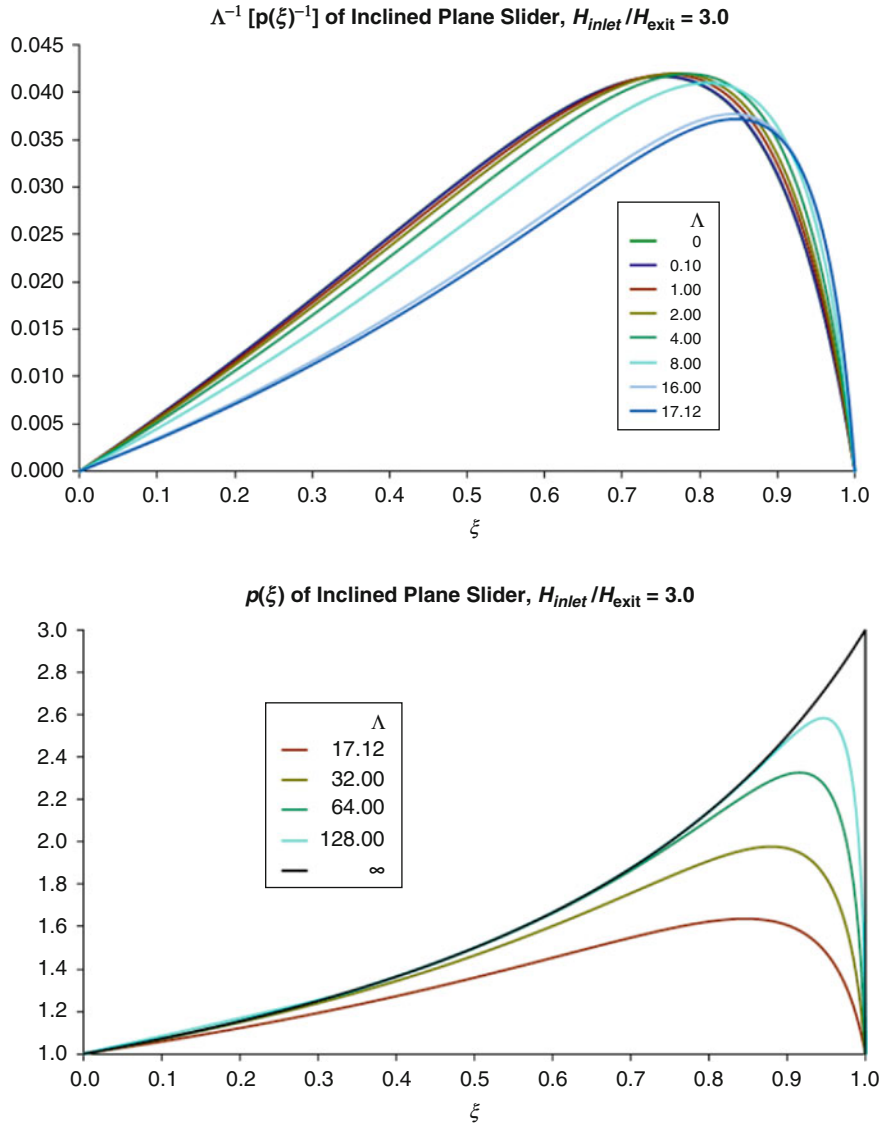
a two-dimensional air-lubricated slider for magnetic recording, Jäger (1987). It should also be possible to use the same procedure to treat the gas-lubricated journal bearing at combinations of Λ, ε and L/D , combinations of interest in MEMS applications.

Key Applications

Power Conversion Equipment, Petrol Chemical Processing Equipment, Machine Tools, Propulsion Engines, Industrial Motors and Generators, Household Appliances, MEMS

Nomenclature

Roman letters	
C	Nominal clearance, m
C_p	Constant pressure specific heat
D	Journal bearing diameter, m
g	Generic gravity constant, m/s^2
\bar{G}	Body force density, m/s^2
h	Film thickness, m
h_i, h_0	Inlet, exit film thickness, m
h	Enthalpy, J/kg
H	h/C
\bar{i}_s	Unit vector along sliding
ℓ	Periphery, width of slider, m



Mathematical Foundation of Fluid Lubrication Theory, Fig. 4 Pressure profile of inclined flat slider for full range of bearing number

L	Journal bearing length, m
p	Absolute pressure, Pa
p_a	Ambient pressure, Pa
p_0	Film pressure at centerline of journal bearing, Pa
P	Normalized film pressure (30), for liquid film, p/p_a for gas film
Q	Flow rate, m^3/s
r	Normalized radial coordinate of thrust bearing

R	Length scale, journal radius, m
\Re	Universal gas constant, $J/kg/K$
t	Time, s
Δt	Time interval, s
\bar{V}_h	Velocity at film height, $= \bar{i}_h V_h$, m/s
V_s	In-plane velocity scale, m/s
\bar{V}	Velocity, m/s

\hat{V}	Rate of deformation tensor, m/s
W	Non-dimensional load capacity of journal bearing
x	Coordinate of slider, m
x, y, z	Cartesian coordinates
Greek letters	
ϕ	Attitude angle of journal bearing, rad
$\bar{\Phi}$	Volume flux, m^2/s
γ	Ratio of specific heats
η_C	Non-Newtonian viscosity factor for Couette flux
η_P	Non-Newtonian viscosity factor for Poiseuille flux
κ	Thermal conductivity, $W/m/K$
\mathcal{A}	Bearing number for gas film, $6\mu V_h(p_a C^2)^{-1} R$
ρ	Mass density, kg/m^3
τ	Non-dimensional time, $= V_s t / C$
$\hat{\tau}$	Stress tensor, Pa
τ_{xy}, \dots	Cartesian matrix element of stress tensor, Pa
$[\tau']$	Diagonal matrix of principal stresses, Pa
$\tau'_I, \tau'_{II}, \tau'_{III}$	Elements of $[\tau']$, Pa
$[\tau'']$	Off-diagonal matrix of shear stresses, Pa
$\tau''_I, \tau''_{II}, \tau''_{III}$	Elements of $[\tau'']$, Pa
ω	Shaft rotation rate, rad/s
ξ	Non-dimensional slider coordinate, x/ℓ
ξ_0	Non-dimensional coordinate of step
Ξ	Non-dimensional pressure flux of gas film
ψ	Sommerfeld angle, (36)
$\bar{\Psi}$	$\bar{\Phi} / (V_s C)$
ζ	Normalized axial coordinate of journal bearing
Mathematical symbols	
$[]$	Square matrix
$\bar{\nabla}$	Gradient operator, m^{-1}
grad	$R \bar{\nabla}$
\times	Vector multiplication
\cdot	Scalar multiplication
$\cdot \cdot$	Scalar invariant of the scalar product of two tensors

Cross-References

- Cavitation Formation and Modeling
- Lubrication with a Grease
- Lubrication with a Non-Newtonian Fluid

References

- G.B. Dubois, F.W. Ocvirk, Analytic derivation and experimental evaluation of short bearing approximation for full journal bearings, NACA Report 1157, 1953.

- L. Gümbel, Verleich der Ergebnisse der rechnerischen Behandlung des Lagerschmierungsproblem mit neueren Versuchsergebnissen. Mbl. Berl. Bez. (VDI), 125-128 (1921)
- W.J. Harrison, The hydrodynamical theory of lubrication with special reference to air as a lubricant. Trans. Camb. Philos. Soc. **xxii** (1912-25), 6-54 (1913)
- D.A. Jäger, An unconditionally stable, high resolution algorithm for gas lubrication problems, Doctoral Dissertation, Mechanical Engineering Department, School of Engineering and Applied Sciences, Columbia University, New York, UMI Order Number 8724042, 1987
- D.A. Jäger, C.H.T. Pan, Function Analysis for the Numerical Treatment of the Theory of Gas Lubrication, in *Mathematics Applied to Fluid Mechanics and Stability: Proceedings of a Conference Dedicated to Richard C. DiPrima*, Rensselaer Polytechnic Institute, Troy, New York, 9-11 Sept 1985, ed. by D.A. Drew, J.E. Flaherty (SIAM, Philadelphia, 1985)
- A.G.M. Michell, Progress in fluid-film lubrication. Trans. ASME **51**(2), 153-163 (1929)
- C.L.M.H. Navier, Memoire sur les Lois du Mouvement des Fluides. Mem. Acad. R. Sci. **vi**(2), 389-440 (1823)
- C.H.T. Pan, R.A. Ibrahim, Cavitation in a short bearing with pressurized lubricant supply. J. Lubr. Technol. **103**(3), 337-349 (1981)
- C.H.T. Pan, L. San Andrés, The narrow groove analysis revisited, in *Proceedings of WTC2005, World Tribology Congress III*, Washington, DC, 12-16 Sept 2005, Paper No. WTC2005-63803, 2005
- L. Rayleigh, Notes on the theory of lubrication. Philos. Mag. J. Sci. **35**, 1-12 (1918)
- O. Reynolds, On the theory of lubrication and its application to Mr. Beauchamp Tower's experiments, including an experimental determination of the viscosity of olive oil. Philos. Trans. R. Soc. Lond. **177**, 157-234 (1886). doi:10.1098/rstl.1886.0005
- A. Sommerfeld, Zur hydrodynamischen Theorie der Schmiermitt-tehreitung. Z. Math. Phys. **50**, 97-155 (1904)
- W. Stieber, *Das Schwimmlager* (Verein Deutscher Ingenieure, Berlin, 1933)
- G.G. Stokes, On the theory of internal friction of fluids in motion. Trans. Camb. Philos. Soc. **viii**, 287-319 (1845)
- H.W. Swift, The stability of lubricating films in journal bearings. Proc. Inst. Civ. Eng. **233**, 267-288 (1932)

Mathematical Fractals

- Fractal Geometry

Maximum Temperature Rise in Sliding

- Flash Temperature Theory

MCM – Machine Condition Monitoring

- [Fault Diagnosis of Machinery Based on an Integrated Approach](#)

Mean Roughness of a Surface (R_a)

- [Surface Statistics and Probability Density Function](#)

Measured Parameters (MP) for Failure Monitoring

- [Journal Bearing Failure Monitoring \(JBFM\)](#)

Measurement of Contact Temperatures

- [Contact Temperature Measurement](#)

Measurement of Flash Temperatures

- [Contact Temperature Measurement](#)

Measurement of Friction-Induced Temperatures

- [Contact Temperature Measurement](#)

Measurement of Interfacial Temperature

- [Contact Temperature Measurement](#)

Measurement of Pressure Distribution

IVAN KRUPKA

Brno University of Technology, Brno, Czech Republic

Synonyms

[Pressure measurement](#)

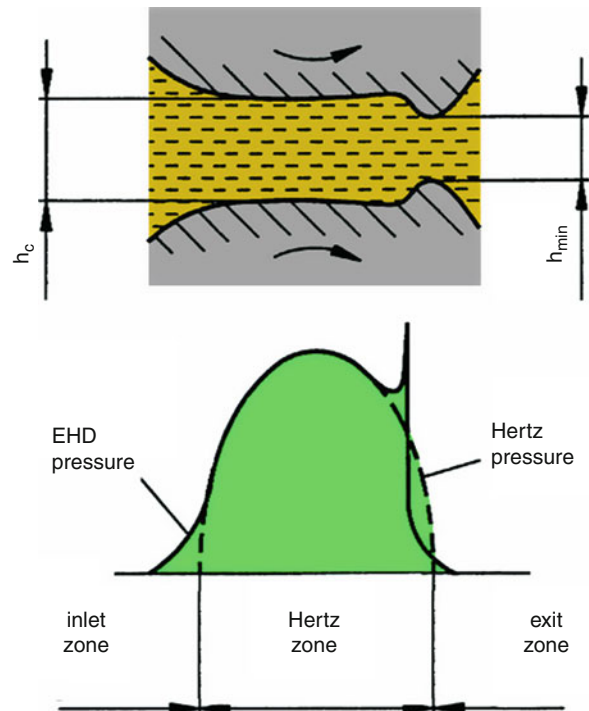
Definition

Evaluation of pressure distribution between non-conformal surfaces by contact or non-contact measurement techniques.

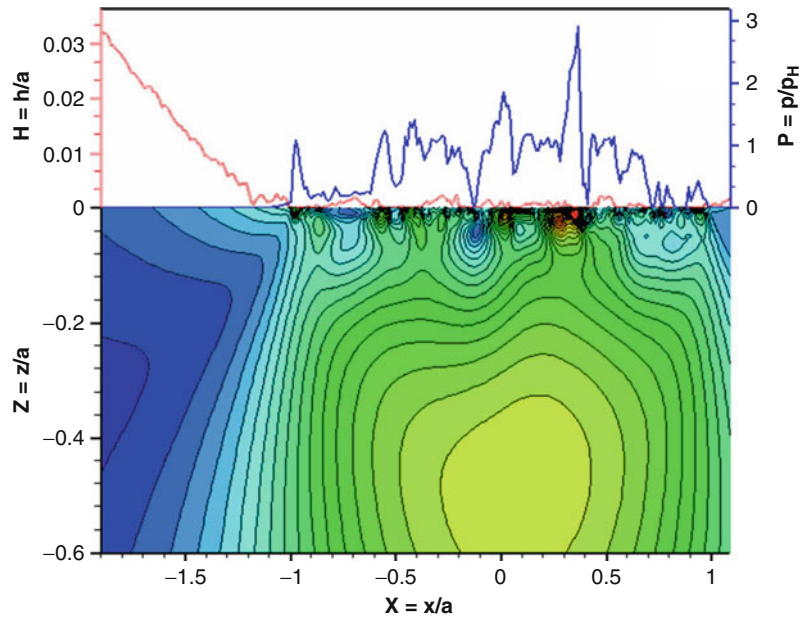
Scientific Fundamentals

Pressure Distribution Within Non-conformal Contacts

Pressure distribution measurement within non-conformal contacts is limited by the fact that the contact area is



Measurement of Pressure Distribution, Fig. 1 Characteristic features of pressure distribution between smooth non-conformal surfaces



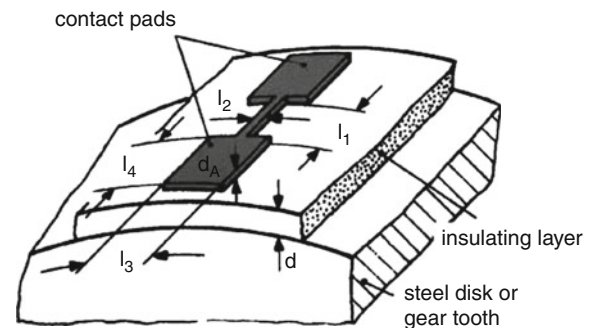
Measurement of Pressure Distribution, Fig. 2 Effect of surface roughness on pressure fluctuations and subsurface stresses within mixed lubricated contact (Zhu et al. 2007)

usually very small and pressure can change over the contact area significantly. Figure 1 shows a typical elastohydrodynamic (EHD) film profile between smooth, non-conformal surfaces (in the direction of motion).

There is a flat zone in the central (Hertz) area and an exit constriction. The pressure distribution is close to that predicted by the Hertz theory, with some deviations in the inlet and exit zone. The pressure (Petrusevich) spike near the exit zone can produce large shear stress that can reduce the fatigue life of machine components. The spatial resolution of the measurement technique should be sufficient enough to map these highly localized pressure changes. Similar requirements can be addressed to the evaluation of pressure around surface asperities within mixed lubricated contact. Figure 2 shows how the presence of rough ground surface within the lubricated contact influences pressure distribution. Asperities' interactions significantly increase pressure and consequently subsurface stress. That is why there is a continuous effort to develop measurement techniques that provide detailed pressure distribution with satisfactory spatial resolution.

Measurement Techniques

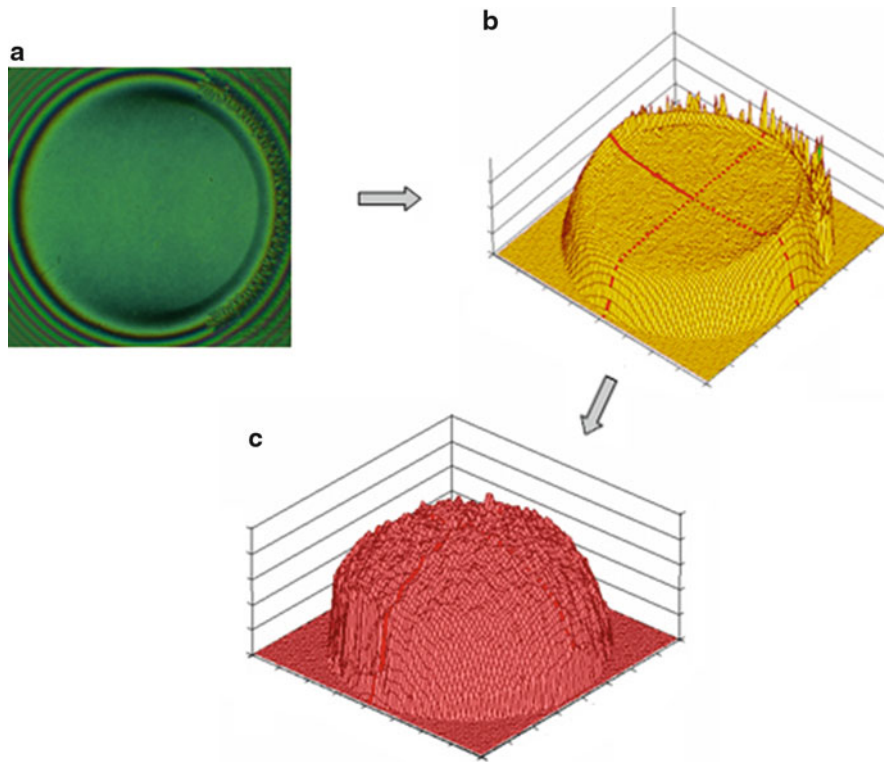
Several experimental approaches have been developed in the effort to obtain pressure distribution within lubricated, non-conformal contacts. These include the



$l_1 = 1.5 \text{ mm}$ $d_A = 0.1 \mu\text{m}$
 $l_2 = 5\text{--}35 \mu\text{m}$ $d = 1\text{--}2 \mu\text{m}$
 $l_3 = 3 \text{ mm}$
 $l_4 = 6\text{--}8 \text{ mm}$

Measurement of Pressure Distribution, Fig. 3 Schematic view on pressure sensor and its dimensions (Höhn et al. 2006)

use of ultrasound detection (Marshall et al. 2004), pressure-sensitive surface transducers (Kannell et al. 1965), Raman spectroscopy (Gardiner et al. 1983), and laser profilometry (Diaconescu and Glovnea 2006). Spatial resolution and evaluation time are the most limiting parameters to obtaining pressure distribution within

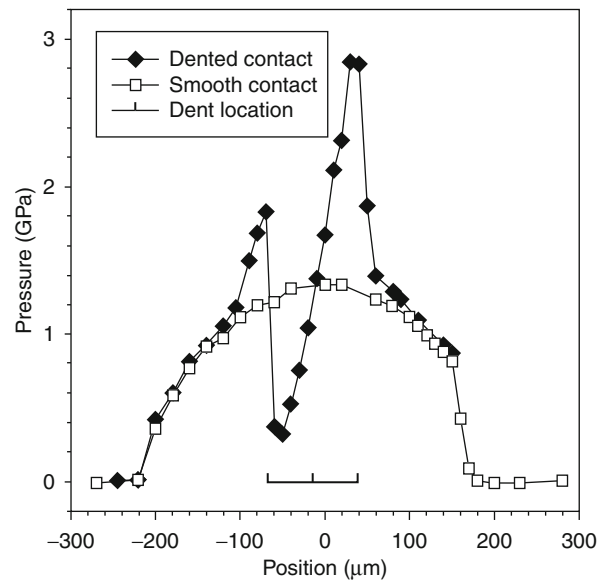


Measurement of Pressure Distribution, Fig. 4 Evaluation of EHD pressure distribution (c) from measured film thickness (b) obtained from chromatic interferogram (a)

lubricated contacts between non-smooth surfaces or contacts operated under non-steady-state conditions.

A technique based on ultrasonic detection uses the reflection of ultrasound at the interface between contacting bodies. A map of reflected ultrasound is obtained by scanning the contact area and it is used to determine the stiffness of the interface that correlates qualitatively with contact pressure. Real engineering surfaces and materials can be used; however, this technique is not suited to the measurement of local transient pressures because of poor spatial resolution and the time required for scanning the contact area.

Pressure-sensitive surface transducers enable the evaluation of pressure distribution within non-conformal contacts through the detection of changes in sensor resistance. The sensor exposed to pressure increases its resistance, which is measured in a Wheatstone bridge. Figure 3 shows a schematic view of a surface with a pressure sensor. The base material of the steel disk is covered by an isolating layer made of aluminum oxide, and the sensor on top is made of Manganin (CuMn12Ni).



Measurement of Pressure Distribution, Fig. 5 Effect of surface dent on EHD pressure distribution (Coulon et al. 2004)

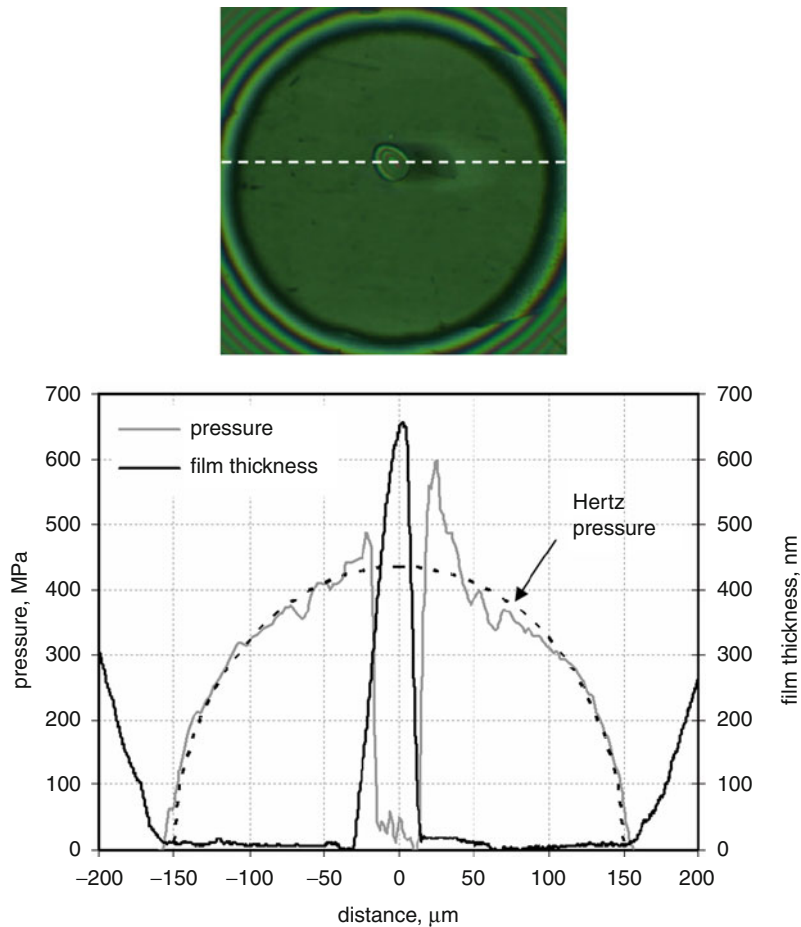
The advantage of this method is that the measurements are made in a metal/metal contact. Conversely, its main limitations are relatively poor spatial resolution and the fact that rubbing surfaces must be separated by thicker lubrication film in order to protect the sensors against mechanical damage.

Other techniques use the detection of changes in the optical properties of lubrication film to evaluate the pressure distribution. They require a model contact in which one surface is made of transparent material (glass or sapphire). Lubrication film subjected to high contact pressure changes its optical properties in the ultraviolet, visible, or infrared of the electromagnetic spectrum that can be recorded by an appropriate sensor.

Raman spectroscopy has been successfully used to map pressure within lubricated dented contact

(Coulon et al. 2004). Nevertheless, it is a scanning method and it takes some time to obtain the pressure distribution within the contact. This limits its use to conditions where the smooth surface is moving over a stationary dented surface. Moreover, this technique is not suitable for thin lubrication films as the intensity of the scattered signal of the lubricant in EHD contacts is weak and decreases with film thickness.

With the development of more precise film thickness measurement techniques, attention has been focused on the evaluation of contact pressure from measured film thickness. This is because the contact pressure is in relation with deformation of rubbing surfaces, which can be derived from experimentally evaluated film thickness and the determination of contact pressure from surface deformations is an inverse problem based on the elasticity



Measurement of Pressure Distribution, Fig. 6 Pressure profile under very thin lubrication film conditions for a dent, 73 μm diameter, 700 nm deep (Krupka et al. 2009)

theory (Åström and Venner 1994). This approach can provide pressure distribution even under transient operational conditions with spatial resolution below $1\text{ }\mu\text{m}$. Its principle is depicted in Fig. 4. Even though the pressure distribution evaluated from measured film thickness is very sensitive to the accuracy of evaluated film thickness, this approach still represents one of the most promising tools for EHD pressure determination due to the relatively wide availability of film thickness measurement techniques. Chromatic interferograms can be captured within a few microseconds so that this approach can be used to obtain pressure distribution even within non-smooth, lubricated contacts operated under very thin film and/or transient conditions.

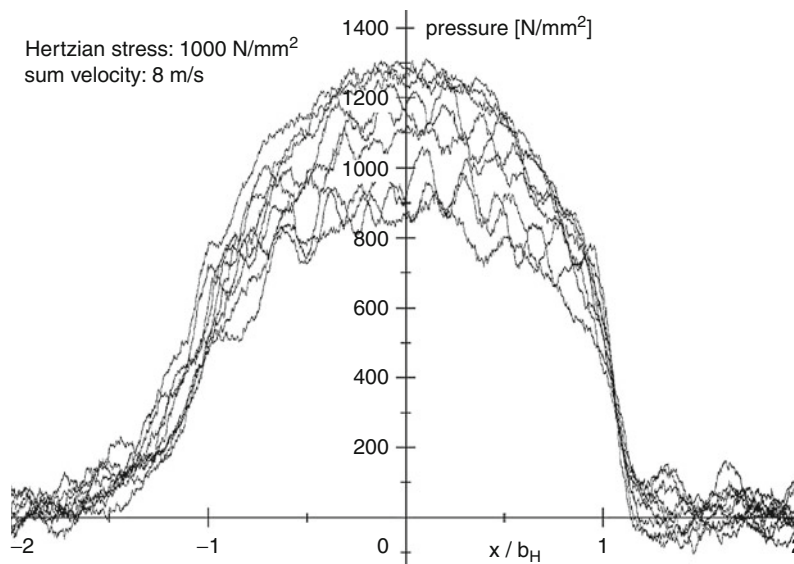
A different approach based on measured film thickness has been developed by Cann and Spikes (2005). The contact pressure is evaluated from deformation of a silica layer that is evaporated on the transparent disk surface that is in contact with the steel ball. The thin film optical interferometry technique is used to measure the spacer-layer thickness and the compression that occurs under loading conditions. Obtained data are compared with the calibration table where the silica thickness is plotted against the contact pressure calculated using the Hertz theory. Nevertheless, this approach enables obtaining pressure distribution within non-smooth, dry static contacts only.

Key Applications

EHD Lubrication of Dented Surfaces

Fatigue life of machine components (e.g., rolling bearings and gears) has increased significantly in the past several decades due to constant improvements in material quality and manufacturing process. Rolling contact fatigue (RCF) initiated from the surface, rather than from inclusion, has become one of the most significant failure modes of machine components. Surface-initiated failure can be caused by the contamination of lubricant even prior to the first use and further contaminants are produced by running-in or wear of rubbing surfaces. The typical size of particles (up to $50\text{ }\mu\text{m}$) is generally large in comparison with lubricant film thickness (less than $1\text{ }\mu\text{m}$). When debris or wear particles pass through the contact, they generate dents on the rubbing surfaces. The indented rubbing surfaces are then subjected to repeated over-rolling as the machine component operates and dents function as stress raisers to initiate microcracks on the surface that grow into spalls, leading to fatigue failure. That is why most attention has been devoted to the evaluation of pressure distribution within dented contacts.

Figure 5 shows the effect of surface dents on pressure distribution within sliding EHD contact observed with Raman spectroscopy. The presence of a dent having several micrometers in depth significantly modifies the



Measurement of Pressure Distribution, Fig. 7 Pressure fluctuation within the contact between rough surfaces (Höhn et al. 2006)

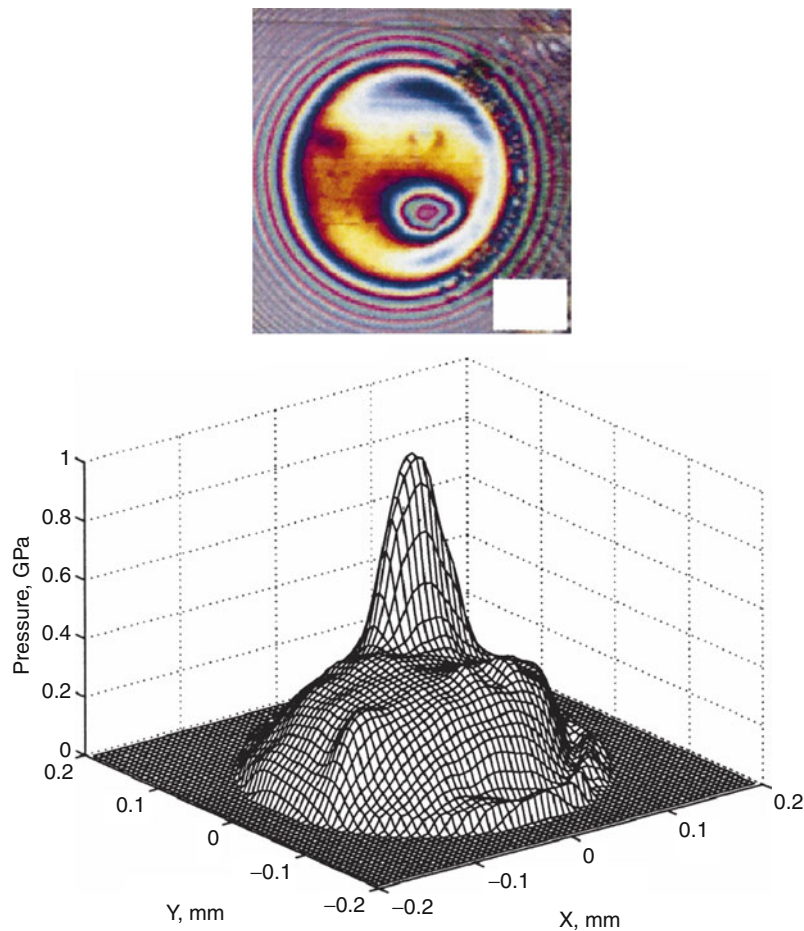
pressure field. The pressure fluctuations are localized around the dent perimeter where the lip of raised metal is present. There are two pressure peaks (1.88 and 2.85 GPa). The first is located in the area of the left dent shoulder and the second peak position corresponds to a minimum film thickness generated by the right dent shoulder. Inside the dent, pressure collapses down to a very low value (0.37 GPa).

Figure 6 shows a chromatic interferogram and film thickness and pressure profiles in the direction of motion taken through the dent that depicts the effect of a dent on EHD contact operated under very thin lubrication film conditions. The smooth glass disk is moving faster than the dented ball so that lubricant is emitted downstream of the dent. The pressure distribution under thin film lubrication conditions is close to Hertzian pressure distribution. Nevertheless, one can notice a significant increase in

pressure distribution downstream of the leading edge caused by the elastic deformation of rubbing surfaces by the lubricant leaving the dent.

EHD Lubrication of Rough Surfaces

Rubbing surfaces of machine elements contain a large number of roughness features of various sizes and shapes produced either during surface finishing procedures or their operation. These roughness irregularities can significantly increase contact pressure and consequently subsurface stresses within concentrated contacts. The knowledge of pressure distribution is a key factor to the understanding of processes taking place between rough surfaces. However, currently no measurement technique mentioned above is capable of providing such a distribution under mixed lubrication conditions. Most limitations arise from low spatial resolution and scanning



Measurement of Pressure Distribution, Fig. 8 Effect of soap lump on pressure distribution within grease lubricated EHD contact (Larsson et al. 2000)

procedure. Pressure-sensitive surface transducers enable mapping of EHD pressure fluctuations caused by surface features (Fig. 7); however, their use under mixed lubrication conditions is not realistic due to their mechanical damage.

Techniques based on the evaluation of contact pressure from measured film thickness can provide sufficient spatial resolution to map the changes in film thickness even in the vicinity of real roughness features. However, undeformed surface topography has to be subtracted from measured film thickness data (to obtain the deformation and consequently the pressure), which has not been successfully solved for real rough surfaces by now.

Grease Lubrication

Greases, consisting of base oil and a thickener, are widely used in heavily loaded machine components such as rolling bearings and gears. They are useful in inaccessible parts where the supply of lubricant cannot easily be renewed and they resist being squeezed away. They also act as a seal to prevent the entry of contaminants and water. However, it has been observed that soap thickener has a significant effect on the lubrication (Larsson et al. 2000). The soap lumps can cause deep elastic indentations accompanied by large pressure fluctuations (Fig. 8) that can influence the life of the lubricated components.

Cross-References

- [EHL Film Thickness Behavior](#)
- [Lubrication with a Grease](#)
- [Point Contact EHL](#)

References

- H. Åström, C.H. Venner, Soap-thickener induced local pressure fluctuations in a grease lubricated elastohydrodynamic point contact. *Proc. IMechE. J. Eng. Tribol.* **208**, 191–198 (1994)
- P.M. Cann, H.A. Spikes, Measurement of pressure distribution in EHL-development of method and application to dry static contacts. *Tribol. Trans.* **48**, 474–483 (2005)
- S. Coulon et al., Pressure profiles measured within lubricated contacts in presence of dented surfaces. Comparison with numerical models. *Tribol. Int.* **37**, 111–117 (2004)
- E. Diaconescu, M. Glovnea, Visualization and measurement of contact area by reflectivity. *ASME J. Tribol.* **128**, 915–917 (2006)
- D.J. Gardiner et al., Raman spectra of lubricants in elastohydrodynamic entrappings. *Wear* **91**, 111–114 (1983)
- B.R. Höhn et al., Influence of surface roughness on pressure distribution and film thickness in EHL contacts. *Tribol. Int.* **39**, 1719–1725 (2006)
- J.W. Kannell et al., Methods for determining pressure distributions in lubricated contacts. *ASLE Trans.* **8**, 250–270 (1965)
- I. Krupka et al., Effect of surface dents on contact pressure in EHD contacts. *Proc. IMechE., J. Eng. Tribol.*, **223**, 683–693 (2009)
- P.O. Larsson et al., Pressure fluctuations as grease soaps pass through an EHL contact. *Tribol. Int.* **33**, 211–216 (2000)
- M.B. Marshall et al., An ultrasonic approach for contact stress mapping in machine joints and concentrated contacts. *J. Strain Anal. Eng. Des.* **39**, 339–350 (2004)
- D. Zhu et al., Simulation of sliding wear in mixed lubrication. *ASME J. Tribol.* **129**, 544–552 (2007)

Measurement of Rubbing Temperatures

- [Contact Temperature Measurement](#)

Mechanical Face Seal

- [Mechanical Seal Analysis](#)

Mechanical Properties After Induction Heat Treatment

- [Induction Heat Treatment and Tribological Property Modification](#)

Mechanical Properties Related to Fatigue

- [Fatigue Limit](#)

Mechanical Pumps

- [Compressor Oils](#)

Mechanical Seal

- [Mechanical Seal Analysis](#)

Mechanical Seal Analysis

ALAN O. LEBECK

Mechanical Seal Technology Inc, Albuquerque, NM, USA

Synonyms

Face seal analysis or modeling; Mechanical face seal; Mechanical seal

Definition

Mechanical seal analysis is the application of basic engineering and tribological principles to predict the performance (leakage, wear, heat generation) of a mechanical face seal. The set of principles and analysis methods used can be called a seal model. Mechanical seal analysis is commonly used as a part of a seal design process to guide the selection of many design parameters based on predicted results. It is also commonly used to understand and diagnose existing designs.

Scientific Fundamentals

Face Seal Fundamentals

Figure 1 illustrates the functions in a mechanical face seal. A chamber contains the sealed fluid, usually under pressure. A shaft passes into the chamber, and the fluid is held in the chamber by a mechanical seal, as shown. The purpose for the seal is to minimize or control the amount of

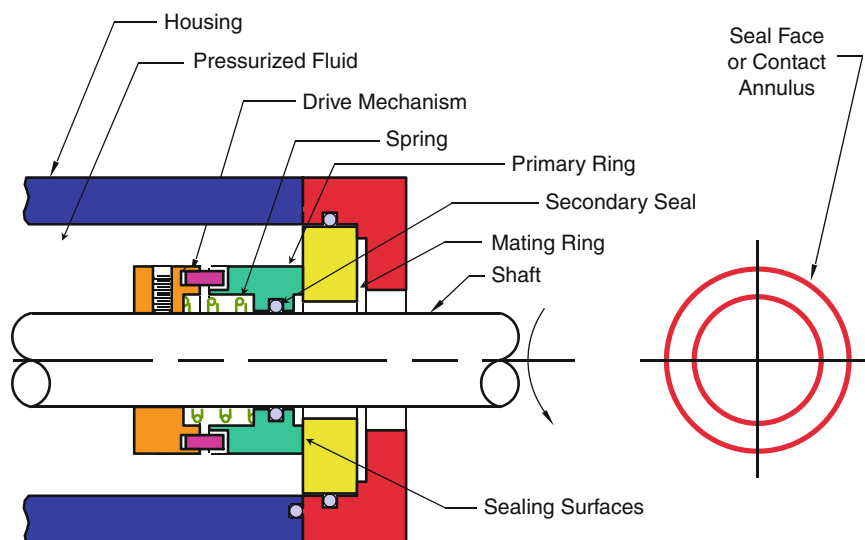
fluid leakage. The sealed fluid attempts to leak between the two seal faces formed by the mating ring (rigidly mounted) and the primary ring (flexibly mounted). These faces relatively tangentially slide as the shaft rotates. The primary face ring has a spring (or equivalent) pushing it toward the mating ring face. It is flexibly mounted, such as by an o-ring, so that it can axially and angularly move so as to be freely pushed into contact (or near contact) and be aligned by the mating face. A flexible rotational drive device connects the primary face to the housing or shaft.

The amount of fluid that flows between the two face surfaces (leakage) is determined by the effective gap between the faces as well as the pressure and viscosity of the fluid. The effective gap is often determined by the roughness of the faces as well as the flatness of the surfaces. In some designs the faces are held apart by fluid pressure alone so as not to contact.

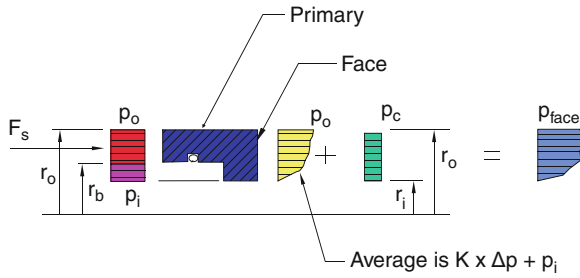
Face Seal Analysis Fundamentals

At a first level of analysis, mechanical seals can be treated as axisymmetric bodies so one need look only at a one-half cross-section as shown in Fig. 2. On the primary seal ring, outside fluid pressure is acting on the seal ring section from r_o down to the balance radius r_b as shown, and inside pressure acts from there to r_i . Balance ratio is the hydraulic load area of the larger pressure to the face area A_f or,

$$B = \frac{\text{Hydraulic load area}}{A_f} = \frac{\pi(r_o^2 - r_b^2)}{\pi(r_o^2 - r_i^2)} = \frac{(d_o^2 - d_b^2)}{(d_o^2 - d_i^2)} \quad (1)$$



Mechanical Seal Analysis, Fig. 1 Mechanical face seal



Mechanical Seal Analysis, Fig. 2 Force balance

Fluid pressure and contact pressures act on the seal face. By definition the average value of the fluid pressure on the face is $K\Delta p + p_i$ where Δp is the net sealed pressure $p_o - p_i$. Defining the average contact pressure as p_c and the spring load as F_s , axial force equilibrium requires that,

$$(K\Delta p + p_i + p_c)A_f = p_o\pi(r_o^2 - r_b^2) + p_i\pi(r_b^2 - r_i^2) + F_s \quad (2)$$

Dividing by A_f and using the definition of balance ratio, contact pressure is

$$p_c = (B - K)\Delta p + p_s \quad (3)$$

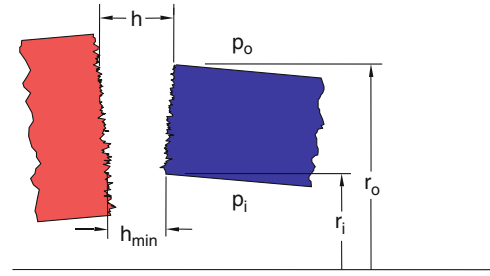
where $p_s = F_s/A_f$

For a common contacting seal p_c must be greater than zero to get an effective seal. The value of K can vary widely (0.3–0.8) and only for well-worn liquid seals approaches 0.5. One may adjust the balance B by selecting the geometrical position of r_b so that p_c is not too small (causing leakage) and not too large (causing excessive wear). For heavily loaded seals B is commonly selected at about 0.8, and for lightly loaded seals where overload is not an issue, B is greater than 1. For non-contacting seals K takes a value so that $p_c = 0$.

Mechanical Seal Analysis

To understand how a seal operates and to provide useful design information, a complete seal model predicts the following:

- Leakage
- Minimum film thickness
- Fluid pressure distribution
- Contact pressure distribution
- Friction heat distribution and total power
- Face temperature distribution
- Deflection of the seal body and the interface shape considering both pressure loads and thermal distortions



Mechanical Seal Analysis, Fig. 3 Film thickness

In a design activity, having such detail from an analysis, one can make precise adjustments in the design to cause seal faces to be essentially parallel and minimize wear and leakage. In a seal diagnostic situation one can often see why the problem is occurring such as when a gas seal is contacting at high speed. To perform such an analysis, a generic method applicable to contacting seals is described here.

To solve for leakage and fluid pressure, consider the seal faces shown in Fig. 3 where the faces are separated by some nominal film thickness $h(r)$, which can be expressed as

$$h(r) = h_{\text{interface}} = h_{\text{interface shape}}(r) + h_{\text{min}} \quad (4)$$

and where the temperature at the interface is assumed to be known as well.

$$T(r) = T_{\text{interface}}(r) \quad (5)$$

For this axisymmetric configuration where the film thickness is small, flow occurs radially across the seal face as pressure caused viscous flow (derivation is same as for classical Reynolds equation in non-sliding direction). At any radius r ,

$$\dot{m} = -2\pi \frac{r\rho h^3}{12\eta} \frac{dp_h}{dr} \quad (6)$$

where p_h is hydrostatic fluid pressure, η is viscosity, ρ is density, and \dot{m} is total mass flow. Recognizing that \dot{m} must be constant at all r , a useful first order differential equation results,

$$\frac{dp_h}{dr} = -\frac{6\eta\dot{m}}{\pi r\rho h^3} \quad (7)$$

Assuming a value for \dot{m} and knowing η , ρ of the fluid ($T_{\text{interface}}$ is known) as well as $h(r)$, (7) can be numerically integrated for pressure from the inside pressure p_i at r_i to the outside radius r_o to find $p_h(r_o)$. Then using a root finding method, the value of \dot{m} that solves the equation,

$$p_h(r_o) - p_o = F(\dot{m}) = 0 \quad (8)$$

is found. One now has $p_h(r)$ matching the boundary values and \dot{m} the leakage.

To solve for contact pressure, actual surfaces have roughness as shown in Fig. 4. These roughnesses can be characterized by height distributions. When contact occurs there is some small overlap of the peak asperities. Defining H as total film thickness between surfaces and h (as used previously) as film thickness between mean surfaces, from Fig. 4,

$$H = z_2 - z_1 + h = z \quad (9)$$

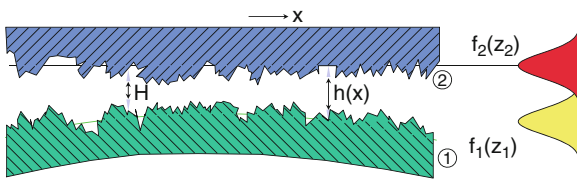
Assuming a simple contact model where S_c is the stress at which an asperity deforms, then

$$p_c = 0, H > 0 \quad (10)$$

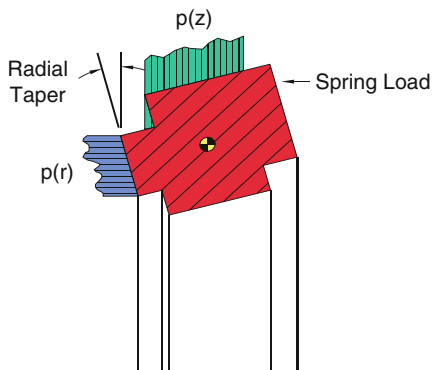
$$p_c = S_c, H \leq 0 \quad (11)$$

One can see that the average value of p_c is just the probability of $H < 0$ times S_c ,

$$\begin{aligned} p_c(r) &= S_c P(H < 0) = S_c \int_{-\infty}^0 f(H) dH \\ &= S_c \int_h^{\infty} f(z) dz = f(h(r)) \end{aligned} \quad (12)$$



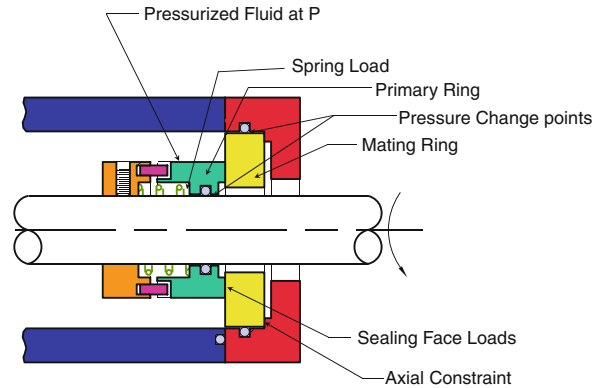
Mechanical Seal Analysis, Fig. 4 Surface roughness



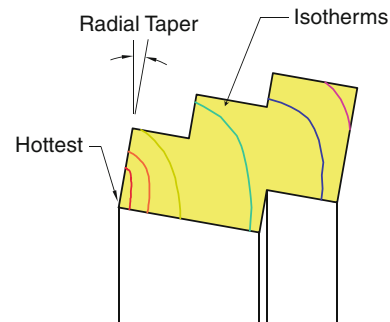
Mechanical Seal Analysis, Fig. 5 Deflection caused by forces

which is evaluated numerically and depends on the probability distribution function f assumed.

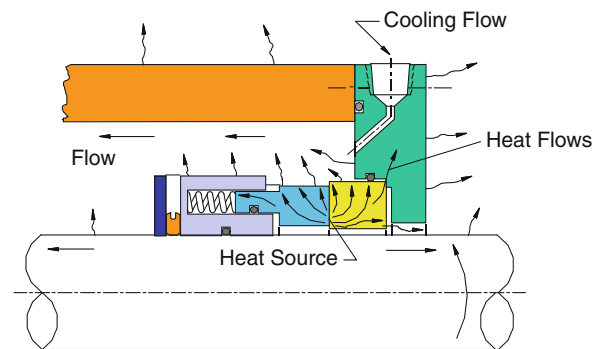
It is important to note from Fig. 4 that $h > 0$ always for real (rough) seal faces, making possible integration of (7).



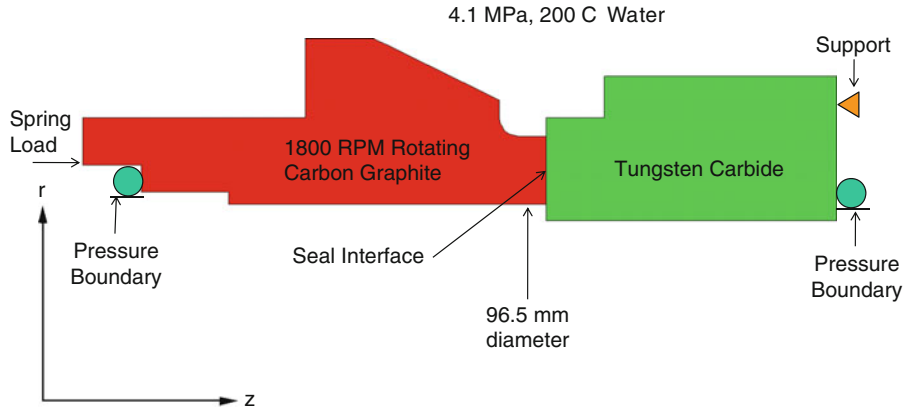
Mechanical Seal Analysis, Fig. 6 Forces and boundary conditions for structural FEA



Mechanical Seal Analysis, Fig. 7 Deflection caused by temperature



Mechanical Seal Analysis, Fig. 8 Boundary conditions for thermal FEA



Mechanical Seal Analysis, Fig. 9 Example problem seal definition

To establish minimum film thickness the integral over the pressures on the face must match the applied load. Based on Fig. 2 and (1) the applied load is

$$W^* = A_f[Bp_o + (1 - B)p_i] + F_s \quad (13)$$

The load support is

$$W = \int_A (p_h + p_c) dA \quad (14)$$

Assuming the interface shape in (4) is constant, load support is a function of minimum film thickness h_{min} . For axial equilibrium,

$$W - W^* = 0 = f(h_{min}) \quad (15)$$

Using iterative methods h_{min} is adjusted to satisfy (15). Now the leakage, minimum film thickness, fluid pressure, and contact pressure are known.

Viscous friction stress τ_f is given by

$$\tau_f = E \left(\frac{\eta U}{H} \right) = \frac{\eta U}{h} = \tau_f(r) \quad (16)$$

where it is assumed that the expected value E of $1/H$ is approximately $1/h$. Contact friction is

$$\tau_c = p_c \mu_c = \tau_c(r) \quad (17)$$

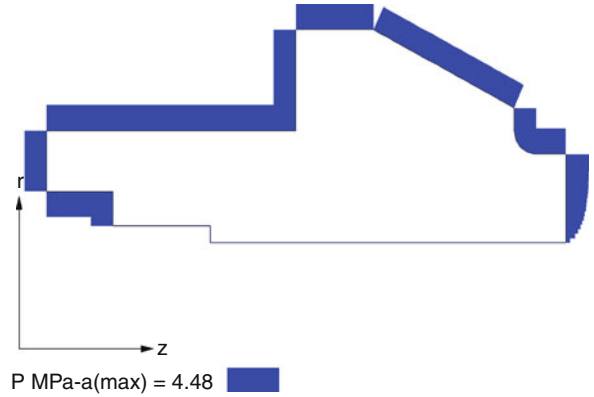
where μ_c is a contact friction coefficient. Heat flux is

$$q = (\tau_f + \tau_c) r \omega \quad (18)$$

And total power P is

$$P = \int_A q(r) dA \quad (19)$$

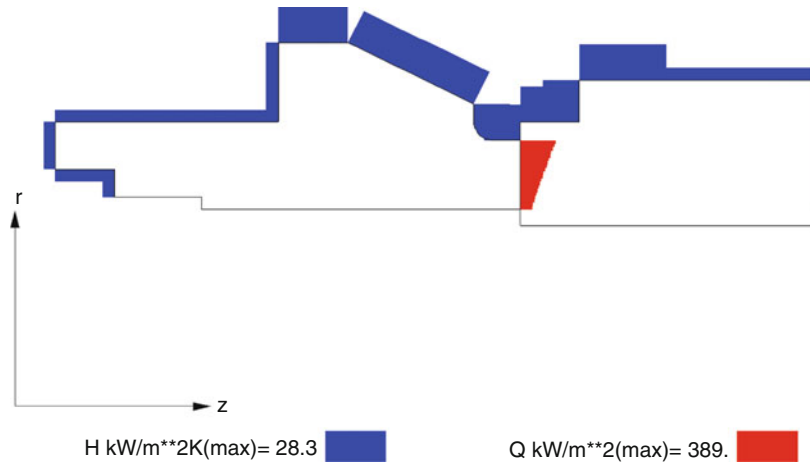
So now face heat distribution and power are known.



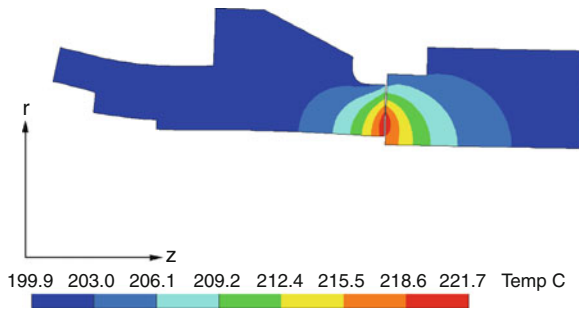
Mechanical Seal Analysis, Fig. 10 Example problem primary pressure boundary

All of the above solutions assumed $h_{interface}$ shape (r) and $T_{interface}$ are known. To find these, finite element methods are generally used. Figure 5 shows how pressure acts on a seal ring to cause distortion of the body and particularly the face. The fluid and contact pressures found above act on the face. Figure 6 shows more generally all of the structural loads on the seal ring that must be included in the finite element structural analysis.

Figure 7 suggests that temperatures that develop in seal rings cause distortion of the face as well. To find these temperatures a finite element analysis of the entire seal system as shown in Fig. 8 must be performed. Here the heat source is the heat flux found above in (18). The temperature is strongly influenced by heat flow from the seal ring boundaries to the surrounding fluid, and these boundary conditions must be included.



Mechanical Seal Analysis, Fig. 11 Example problem heat transfer boundary



Mechanical Seal Analysis, Fig. 12 Example problem predicted temperature and distortion (300x)

In summary, assuming $h_{\text{interface shape}}(r)$ and $T_{\text{interface}}$ one gets $q(r)$, $p_f(r)$, $p_c(r)$ from above. Then one may perform two finite elements analyses,

$$T_{\text{FEA interface}}(r) = \text{FEA}_{\text{thermal}} \begin{bmatrix} q(r) \\ \text{Heat transfer boundary conditions} \\ \text{material properties} \end{bmatrix} \quad (20)$$

$$h_{\text{FEA shape}}(r) = \text{FEA}_{\text{structural}} \begin{bmatrix} p_f(r) \\ p_c(r) \\ \text{Temperature field} \\ \text{Spring load} \\ \text{Pressure boundary conditions} \\ \text{Material properties} \end{bmatrix} \quad (21)$$

At this point an iterative method must be used. Various methods can be used including that of solving large sets of linear equations based on influence

coefficients. When the results from these FEA equal the assumed values or

$$\begin{aligned} h_{\text{interface shape}}(r) &= h_{\text{FEA shape}}(r) \\ T_{\text{interface}}(r) &= T_{\text{FEA interface}}(r) \end{aligned} \quad (22)$$

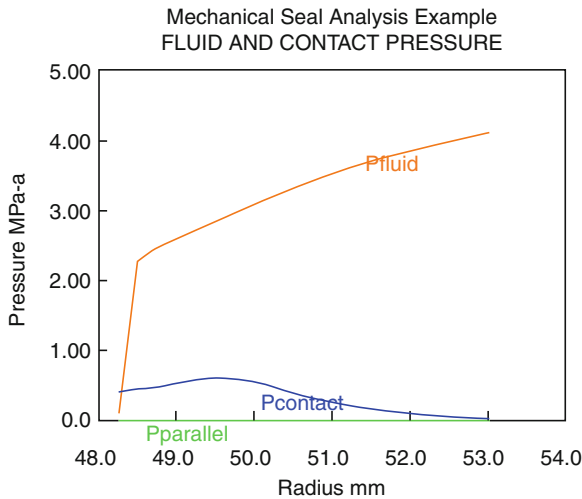
one has a complete equilibrium solution. All of the unknowns of the model are solved.

An example problem is shown in Fig. 9. The pressure boundary is shown in Fig. 10, heat transfer boundary in Fig. 11. Following the methods outlined above, the temperature distribution and distortion predicted are shown in Fig. 12. The actual predicted fluid and contact pressures on the face are shown in Fig. 13 and the predicted interface shape in Fig. 14. Leakage is predicted at 36 g/h, minimum film thickness is 0.48 μm , and friction power is 480 W. The pressure distribution of Fig. 13 shows that the fluid starts out as a liquid at the outside diameter and then a phase change to steam occurs just before reaching the inside diameter. The interface curve shown in Fig. 14 shows both a general tapering of the face plus some curving of the surface.

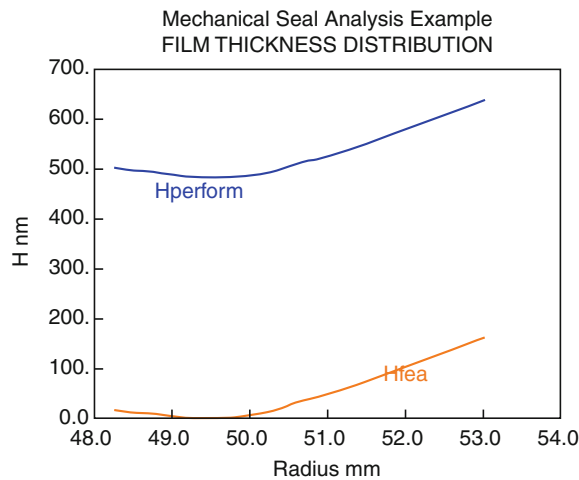
The mechanical seal analysis is complete and the results can be used to understand the seal performance and modify the design.

Literature

The literature on mechanical seals includes many papers that provide some type of mechanical seal analysis. The methods above are described in detail by Lebeck (1991) and Lebeck (1999). Lebeck and Albor (1999) show how mechanical seal analysis similar to above is used to design a gas seal. Ruan et al. (1997) develop a complete model similar to that described above for contacting seals using



Mechanical Seal Analysis, Fig. 13 Example problem predicted fluid and contact pressure



Mechanical Seal Analysis, Fig. 14 Example problem predicted interface shape (Hperform)

influence coefficients based on FEA. A complete solution is found and favorably compared to experiment.

For non-contacting seals that obtain liftoff from the hydrostatic pressure that develops due to radial taper or tilt, there are several models. Etsion and Sharoni (1980) develop analytical solutions for a face seal with radial taper and tilt. Pascovici and Etsion (1992) explore thermal solutions for face seals and how tilt develops. Brunetiere et al. (2003a) develop a semi-analytical model for non-contacting, hydrostatic lift-off type of seals that

develop taper due to thermal heating. Brunetiere et al. (2003b, c) develop a comprehensive thermo elastic hydrodynamic model involving numerical solution for temperatures and thermal distortion and compare predictions to experiment. Most recently Brunetiere and Apostolescu (2009) develop a semi-analytical model for non-contacting, initially radially tapered, hydrostatic lift-off type of seals. Good agreement with experiment is shown.

Polycarpou and Etsion (1998) describe a model to predict leakage in contact seals for gas and consider contact mechanics and slip flow. Comparison to experiment is made.

Key Applications

Mechanical seal analysis is used to understand seal behavior. It is used to diagnose problems such as edge contact. It is commonly used for contact-type seals to adjust a design so that near parallel face operation is achieved. It is used for gas seal design to ensure that there is no contact.

Mechanical seal analysis is used to design seals for all types of pumps and gas seals for compressors. It is also used for ship and submarine shaft seal design.

Cross-References

- ▶ [Finite Difference Method for Fluid-Film Bearings](#)
- ▶ [Friction Coefficient](#)
- ▶ [Hydrodynamic Lubrication](#)
- ▶ [Hydrostatic Thrust Bearings](#)
- ▶ [Lubrication Regimes](#)
- ▶ [Mechanical Seals](#)

References

- N. Brunetiere, A. Apostolescu, A simple approach to the thermoelastohydrodynamic behavior of mechanical face seals. *Tribol. Trans.* **51**, 243–255 (2009)
- N. Brunetiere, B. Tournier, J. Frene, A simple and easy to use TEHD model for non contacting liquid face seals. *Tribol. Trans.* **46**(2), 187–192 (2003a)
- N. Brunetiere, B. Tournier, J. Frene, TEHD lubrication of mechanical face seals in stable tracking mode: part 1 - numerical model and experiments. *J. Tribol.* **125**, 608–616 (2003b)
- N. Brunetiere, B. Tournier, J. Frene, TEHD lubrication of mechanical face seals in stable tracking mode: part 2 - parametric study. *J. Tribol.* **125**, 617–627 (2003c)
- I. Etsion, A. Sharoni, Performance of end-face seals with diametral tilt and coning-hydrostatic effects. *ASLE Trans.* **23**(3), 279–288 (1980)
- A.O. Lebeck, *Principles and design of mechanical face seals* (Wiley, New York, 1991)
- A.O. Lebeck, Mixed lubrication in mechanical face seals with plain faces. *Proc. Inst. Mech. Engrs. Part J* **213**, 163–175 (1999)
- A.O. Lebeck, G. Albor, A double gas seal with coplanar coaxial rayleigh pad faces for pump sealing applications, in *Proceedings, 15th International Pump Users Symposium* (March 1999)

- M.D. Pascovici, I. Etsion, A thermo-hydrodynamic analysis of a mechanical face seal. *ASME J. Tribol.* **114**(4), 639–645 (1992)
- A. Polycarpou, I. Etsion, Static sealing performance of gas mechanical seals including surface roughness and rarefaction effects. *Tribol. Trans.* **41**(4), 531–536 (1998)
- B. Ruan, R.F. Salant, I. Green, A mixed lubrication model of liquid/gas mechanical face seals. *Tribol. Trans.* **40**(4), 647–657 (1997)

Mechanical Seal Face Materials

► Materials for Mechanical Seals

Mechanical Seals

JAMES P. NETZEL
Yorkville, IL, USA

Synonyms

Active lift seal; End face seal; Face seal; Gas lubricated seal; Liquid lubricated seal; Seal; Shaft seal methods; Upstream pumping seal

Mechanical Seals

Definition

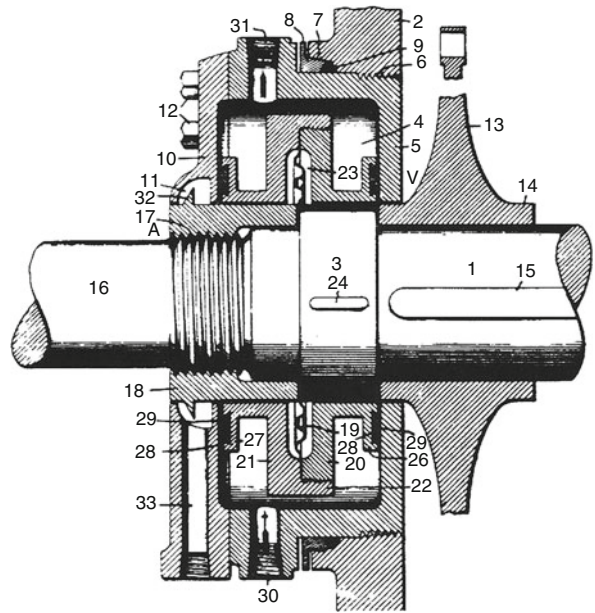
A mechanical seal is a component of a machine to prevent leakage of a fluid along a rotating shaft to atmosphere. Sealing takes place in a plane perpendicular to the shaft.

Introduction

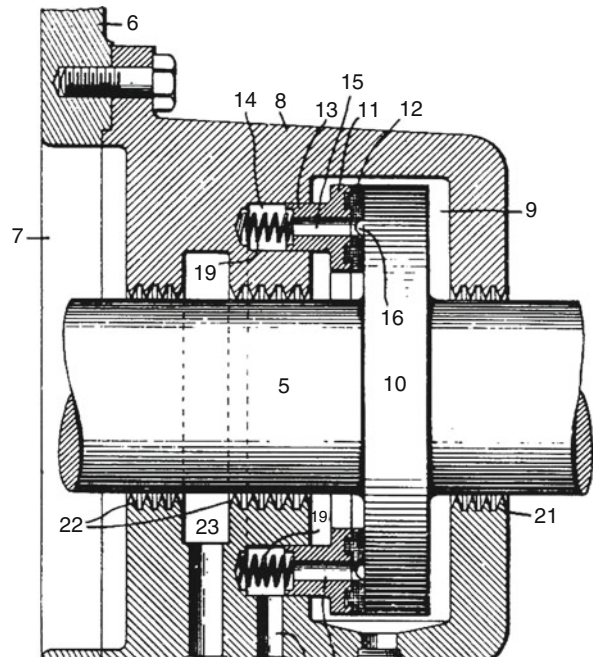
Early in the twentieth century, packing was still the only method of sealing a shaft on rotating equipment. During this time, engineers and designers began looking for a more positive sealing device. The result of their early work is shown in Fig. 1. This design is documented in U.S. Patent 1,063,633 by J. Wilkinson, dated June 13, 1913. The primary seal face in this design is made of carbon.

The inventor referred to this design as a double seal. The inboard seal runs against the housing, item 5. The outboard seal runs against the end plate, item 10. Water is provided between the seals to cool and lubricate the seal faces.

The continued early development in seal design is illustrated in Fig. 2. U.S. Patent 1,315,822 was filed by J.H. Doran on September 9, 1919.



Mechanical Seals, Fig. 1 Early seal design shown in 1913 U.S. Patent



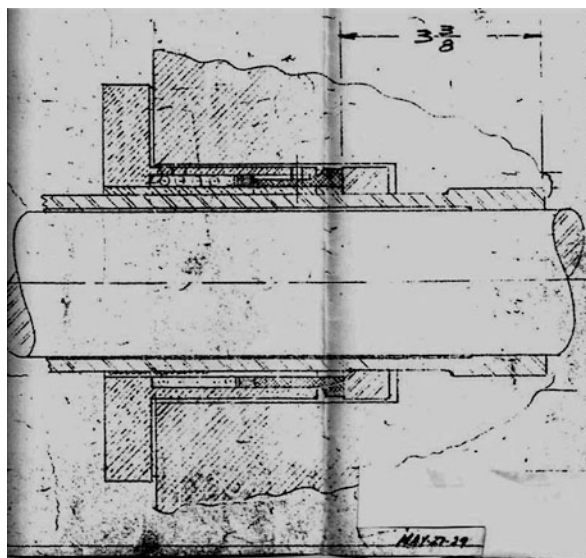
Mechanical Seals, Fig. 2 Patent drawing from 1919 illustrating a single seal

The inventor referred to this design as packing for a steam turbine. However, in present-day terminology this design is a single mechanical seal. Small coil springs, item 19, are used to load the seal face against a shoulder in the shaft, item 10.

By the late 1920s, seal designs were becoming more common. The Ryan seal appeared in May 1929. This was a single seal design. The seal head is held in a stationary position while the mating ring rotates with the shaft. This design is shown in Fig. 3.

This is still a very primitive seal design. A small, square ring of packing is loaded against a bronze seal face by a single coil spring. The bronze seal face runs against a hard steel face. The hard steel face is shrink-fit to a metal sleeve. Since mechanical seals were new to industry, they had to fit the same space as a packing installation. It would take 60 years for pump designers to improve the stuffing box design to provide a better environment for the seal.

Just before and during World War II, significant developments occurred to establish mechanical seals in industry. Improvements in design and materials were occurring at an accelerated rate. The most important development was that of the "O" ring as a static or dynamic seal. A U.S. Patent was granted to N. Christensen in 1937. Today, this component is available in 15 different elastomeric materials. This toroidal-shaped ring revolutionized the seal industry. No longer did the seal face have to run against a structural part of the equipment. Instead, the mating seal face became a replaceable component in the seal assembly.



Mechanical Seals, Fig. 3 Ryan seal for a refinery pump, dated May 1929

The mating ring then became a standardized component. Modern materials of construction became stock items to reduce cost.

By 1948 manufacturing processes were being developed to improve seal performance. The most notable manufacturing change occurred in the development of a precision lapping process for seal faces. This process allows seal faces to be lapped to two light bands flatness or less on a production basis. This was a significant milestone in mechanical seal development, for it allowed a seal to operate at near-zero leakage.

Three significant design improvements also occurred at this time. These were:

1. Seal balance. This was a method to reduce the contact pressure at the seal faces.
2. Full convolution rubber bellows seals.
3. Development of the package or cartridge seal assemblies.

The cartridge seal allowed the seal user to assemble all the seal components as a unit instead of as loose parts.

In the early 1960s, the metal bellows seal was developed for high-temperature refinery service. This development in seal technology extended the temperature for mechanical seals to 800 °F.

The most significant advance in seal technology was the development of spiral groove seal faces. This design feature allowed the application of mechanical seals to high-performance gas compressors. This design is illustrated in Fig. 4.

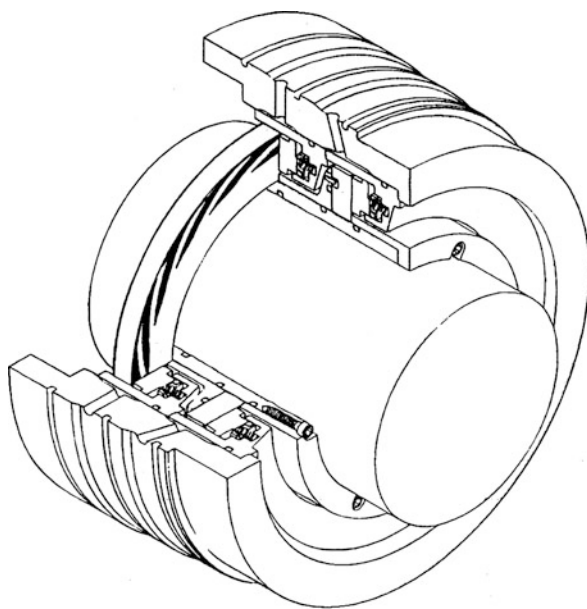
This patented seal design was first applied to gas pipeline equipment in the late 1970s. The design was so successful that it became the industry standard in less than 10 years.

In the early 1990s, this design was extended to pumping applications in chemical and refining plants to eliminate hazardous leakage to the environment.

In the pharmaceutical industry, this sealing technology is used to protect valuable products from conventional liquid lubricants used for contacting seals. In this case, nitrogen or purified air is used to pressurize the seal chamber and act as a seal lubricant.

Gas in the compressor is used to pressurize the seals. When rotated, the spiral grooves in the seal face create a very thin gas film that separates the seal faces. The only heat generated is that of shearing gas at the seal faces. The heat is carried away by a small gas flow through the seal faces and by conduction to the surrounding structural parts of the seals and equipment.

Another significant development that occurred in the early 1990s was that a single seal could operate at the vapor



Mechanical Seals, Fig. 4 Mechanical seal design for high-performance gas compressors (John Crane Inc.)

pressure of the liquid sealed. This design allowed a change of state in the seal faces from a liquid to a gas without damage to the seal from violent flashing of the liquid. When applied to pumps handling cryogenic liquids, seal life was extended from just a few weeks to more than 8 years.

Another development in seal technology occurred in 2005. This concept is referred to as “active lift,” which is an extension of the patented upstream pumping concept. In this case, a very small quantity of liquid is moved from the low-pressure side of the seal to the high-pressure side. The significance of this design is that it promotes full film lubrication. The movement of liquid can be accomplished at very high pressures.

Non-contacting seal technology is the newest development in sealing technology and is already more than 35 years old. Originally developed for high-performance gas compressors, the concept of spiral groove face design has been expanded to other areas of application. This has led to significant understanding of the processes necessary to improve seal and equipment life.

Scientific Fundamentals

Design Features

The basic component parts of a mechanical seal are the primary and mating rings, Fig. 5.

Together, they form the sealing surfaces that are perpendicular to the shaft. The primary ring is a component of the seal head. The secondary seal is a critical component of the seal that forms the seal along the shaft. Metal parts of the seal provide mechanical drive and loading to the seal faces. The mating ring assembly consists of the mating ring and static seal. The seal head rotates with the shaft. For high-speed shafts, the seal head will be stationary while the mating ring assembly will rotate.

The processes occurring at the seal faces are shown in Fig. 6. As the shaft begins to rotate, a fluid film will develop. Frictional heat will also develop for a contacting seal. The heat developed must be limited to prevent the liquid being sealed from flashing or carbonizing. Removal of heat from a contacting seal is accomplished by providing a liquid flush to the seal faces.

Seal Balance

If the hydraulic pressure or load is too high, the excessive heat will result in seal failure. The hydraulic pressure at the seal face can be reduced by what is referred to as balancing. By providing a step in the shaft sleeve or seal hardware, the pressure at the seal faces can be reduced, Fig. 7

The calculation for balance is given in Table 1. Balanced seals can range from 0.65 to 1.00. Unbalanced seals have a balance ratio greater than 1.

Face Pressure

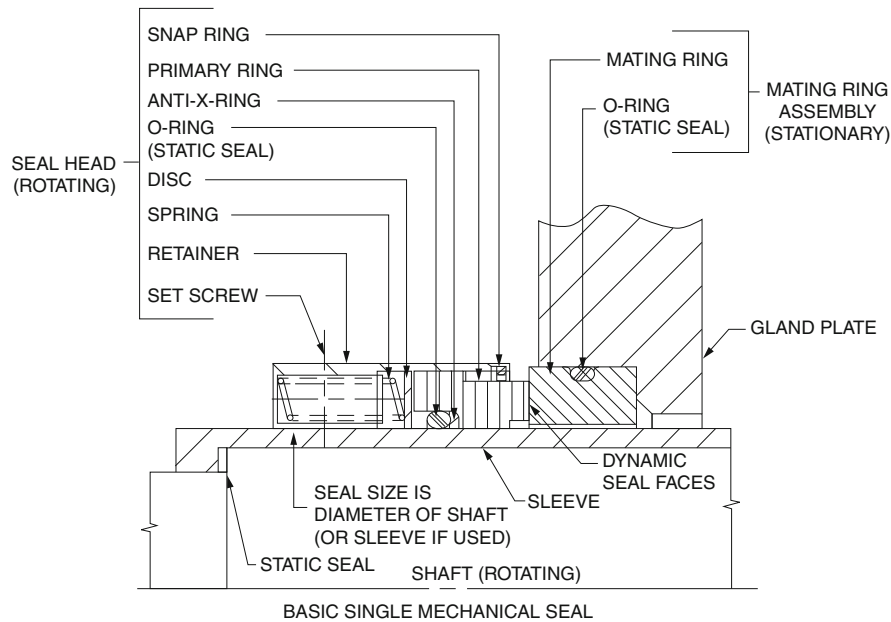
Seal face pressure is an important consideration in seal design. It is the sum of the hydraulic load on the seal and the required mechanical load designed into the seal assembly. The hydraulic load is the product of the pressure differential across seal face multiplied by seal balance minus the pressure gradient, 0.5 (see Table 1).

Pressure-Velocity

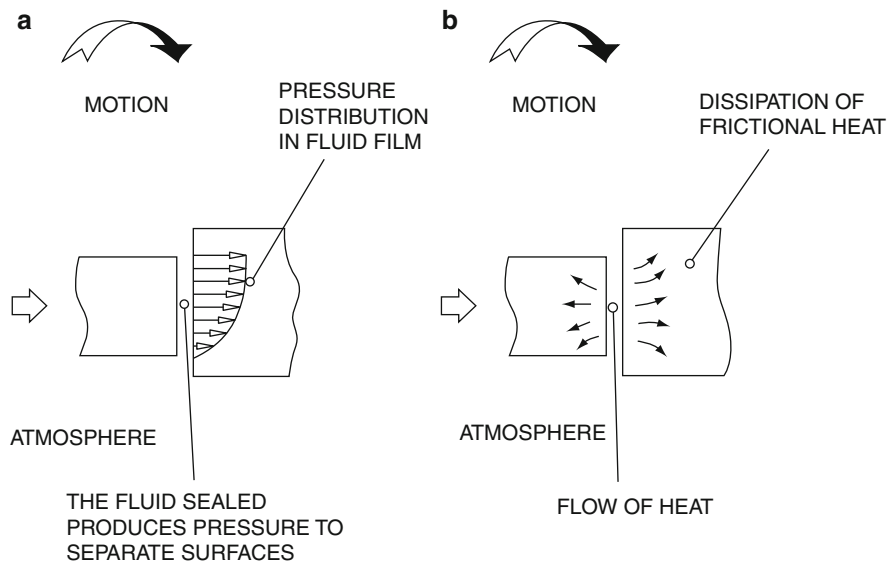
The pressure-velocity (PV) value for a given installation can be valuable when compared with maximum values for materials developed by seal manufacturers. PV is the product of the actual face pressure multiplied by the mean velocity of the seal face, Table 1. A PV value larger than a tested value can result in more wear at the seal faces. PV values for common seal face materials are given in Table 2.

Power Consumption

PV value also enables the estimation of the power loss at the seal faces (see Table 1). For values of the coefficient of friction, see Table 3. Knowing the power loss at the seal and the heat soak from the process, a determination can be made on the correct coolant flow to the seal to prevent damage to the seal faces.



Mechanical Seals, Fig. 5 Basic component parts of a seal



Mechanical Seals, Fig. 6 Processes occurring at the seal faces

Operating Envelope

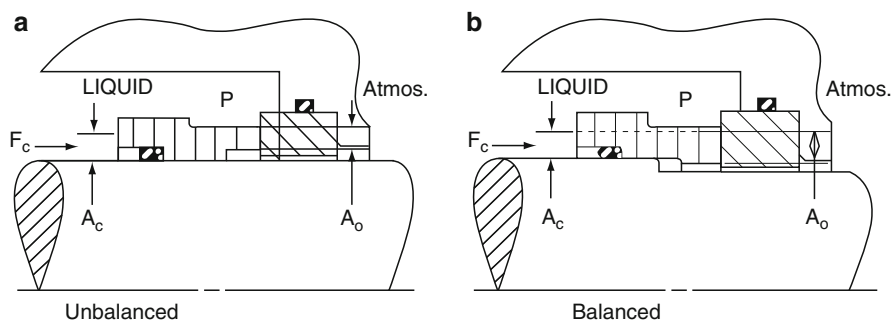
When a seal is applied to a specific liquid it has what is referred to as an operating envelope, Fig. 8. The upper limit of the envelope is determined by wear at the seal faces. This is defined by the materials of construction for the seal faces. The liquid being sealed must be cooled to allow the seal to operate away from the vapor pressure.

Operating within the envelope will result in excellent seal performance.

Seal Face Lubrication

There are two types of lubricating films for seals.

1. Full liquid film
2. Partial liquid/gas film



Mechanical Seals, Fig. 7 Hydraulic pressure acting on the primary ring. (a) Unbalanced, (b) balanced

Mechanical Seals, Table 1 Common seal calculations

Seal balance	
$b = a_c/a_o$	b = balance a_c = Hydraulic area closing a_o = Hydraulic area opening
Face pressure	
$P_f = P_h + P_{sp}$	P_f = Face pressure P_h = Hydraulic pressure P_{sp} = Mechanical pressure for a seal design
$P_h = \Delta P(b - k)$	ΔP = Pressure differential across seal face b = Seal balance k = Pressure gradient factor (normally 0.5)
$P_{sp} = F_{sp}/a_o$	F_{sp} = Seal spring load a_o = Seal face area
Seal face pressure can be expressed as	
$P_f = \Delta P(b - k) + P_{sp}$	
Pressure velocity	
$PV = P_f V_m$	V_m = Velocity at mean seal face diameter
$P_f V_m = [\Delta P(b - k) + P_{sp}] V_m$	
Power consumption	
$N_f = (PV) f a_o$	f = Coefficient of friction

Full liquid film occurs in a contacting seal when the fluid sealed is an oil-based product, Fig. 9. This type of film can also be generated by the design of the seal faces. This is a major advancement in seal design.

A partial liquid/gas film will occur on water-type solutions and petroleum products such as propane. For a seal to be successful, the amount of liquid film on

Mechanical Seals, Table 2 PV limitations for common seal face materials

Sliding materials		"PV" limit	
Rotating	Stationary	bar \times m/s	lbs/in ² \times ft/min
Carbon-graphite	Ni resist	35.03	100,000
	Ceramic	35.03	100,000
	Tungsten carbide	175.2	500,000
	Silicon carbide	175.2	500,000
Tungsten carbide	Tungsten carbide	42.02	120,000
Silicon carbide	Silicon carbide	175.2	500,000
Silicon carbide	UNCD/Silicon carbide	122.6	350,000

(1) UNCD trade mark for Ultrananocrystalline Diamond by Advance Diamond Technologies. (2) Values given are for water services. (3) For limits in oil services multiply by 1.6

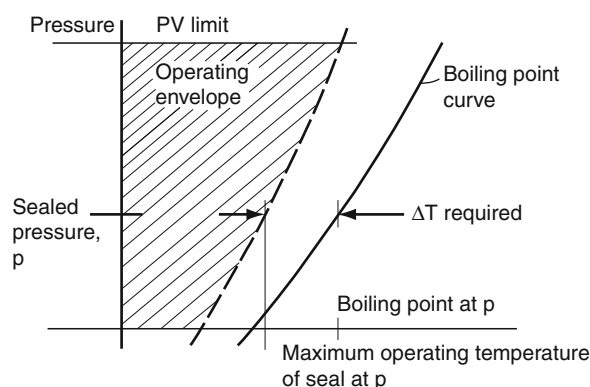
Mechanical Seals, Table 3 Coefficient of friction for seal face materials

Sliding materials		
Rotating	Stationary	Coefficient of friction
Carbon graphite (Resin filled)	Cast iron	0.07
	Ceramic	0.07
	Tungsten carbide	0.07
	Silicon carbide	0.02
Silicon carbide	Tungsten carbide	0.05
Silicon carbide	UNCD/Silicon carbide	0.02–0.05

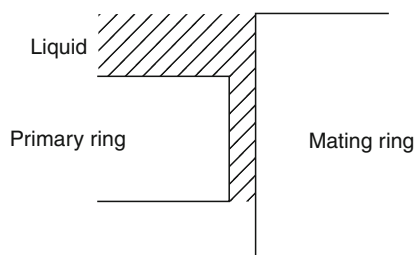
UNCD is a trade name for Ultrananocrystalline Diamond by Advance Diamond Technologies, Inc.

a contacting seal should be greater than 50% of the seal face width (Fig. 10). If there is no lubricating film for a contacting seal, damage and seal failure will occur.

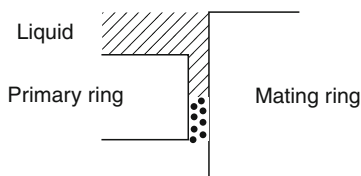
For a seal to achieve maximum service performance, the seal faces must be close to parallel to one another. Angular deflection at the seal faces must be controlled by design. Excessive positive angular deflection will result in unwanted leakage as the seal faces open to the fluid being sealed. Negative angular deflection will result in the heavy contact at the outside diameter seal face, preventing the liquid being sealed from entering the seal faces to promote lubrication.



Mechanical Seals, Fig. 8 Operating envelope for a seal



Mechanical Seals, Fig. 9 Full liquid film



Mechanical Seals, Fig. 10 Partial liquid/gas film

Key Applications

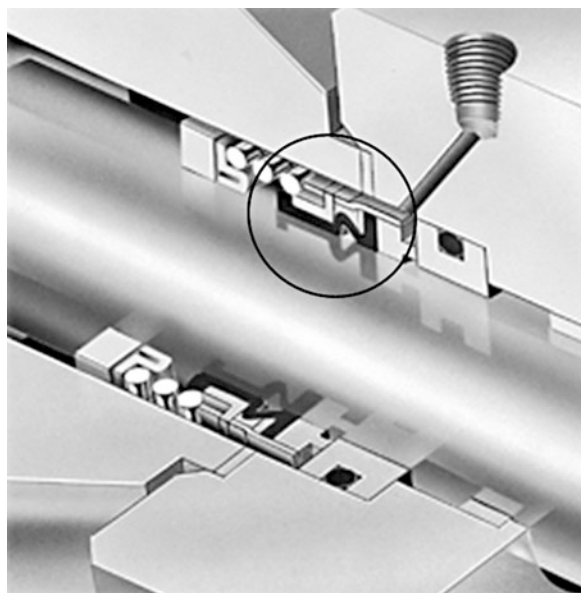
General Service

Today, with over 100 years of advancement in seal technology, seals have become the dominant choice for sealing a rotating shaft. The mechanical seal is a component of a pump or other type of rotating equipment. It must be capable of operating in an almost infinite number of fluids at various operating conditions of pressure, temperature, and shaft speed. Advances in seal design and materials have resulted in exceptional performance and seal life. One design that has been proven over time is the full convolution bellows seal, Fig. 11.

This design has been applied to many applications in industry, including ship board pumps. In this design, the tail of the bellows is held to the shaft by a drive band. This squeeze fit seals the shaft and allows the seal to rotate with the shaft. Static sealing is accomplished at the front of the bellows and at the back of the primary ring. This bellows seal design is more flexible than other seal designs and can tolerate greater shaft motion and runout.

Refinery Service

Sealing requirements in the petroleum refining industry must meet the emission standards defined by the U.S. Clean Air Act of 1990. This has resulted in the development of American Petroleum Institute Standard API 682. This standard was developed by users with input from



Mechanical Seals, Fig. 11 Full convolution bellows seal (John Crane, Inc.)

equipment and seal manufacturers. The goal of this standard was to meet emissions requirements and provide a seal life with a minimum of 3 years. Seals for API services have been verified by a seal manufacturer's test under simulated refinery conditions. These service areas are:

Type A: A single seal mounted inside the seal chamber with a rotating seal head, Fig. 12.

This is a balanced cartridge design with multiple springs and an "O" ring as a secondary seal. This seal is preferred for all refinery services except for non-flashing hydrocarbons above 150°C (300°F). The Type A design is the standard for temperatures up to 265°C (500°F).

Type B: A single, low-temperature metal bellows seal mounted inside the seal chamber. The metal bellows seal head is made of a nickel alloy and rotates with

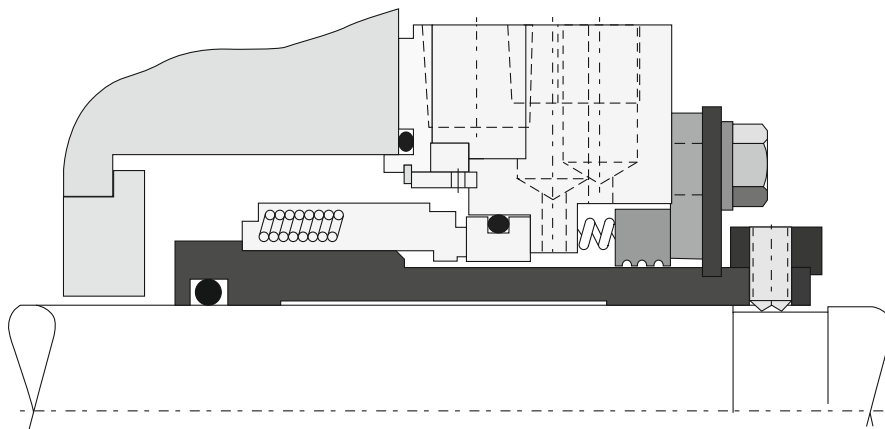
the shaft. "O" rings in the design are made of a fluorocarbon elastomer.

This low-temperature seal design, Fig. 13, is a standard optional selection for non-flashing hydrocarbon services up to 150°C (300°F).

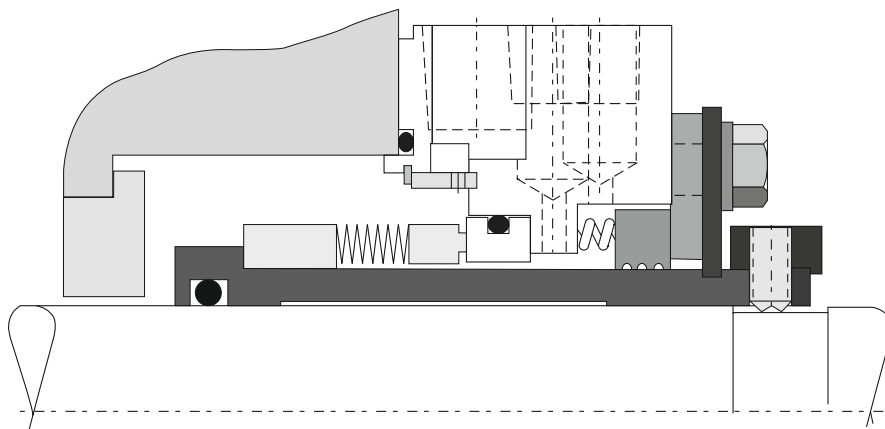
Type C: A single, high-temperature metal bellows seal mounted inside the seal chamber. High-temperature static seals are flexible graphite. This seal is the standard selection for non-flashing hydrocarbon applications when temperatures are above 150°C (300°F) and pressures are less than 17 bar (250 psig), Fig. 14.

Gas-Lubricated Seals

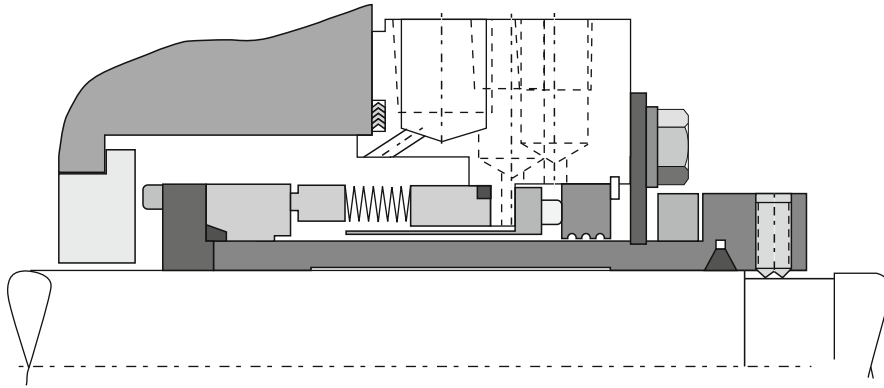
There are many applications in industry that require a double seal. Some of these installations cannot use a liquid barrier to cool and lubricate the seals.



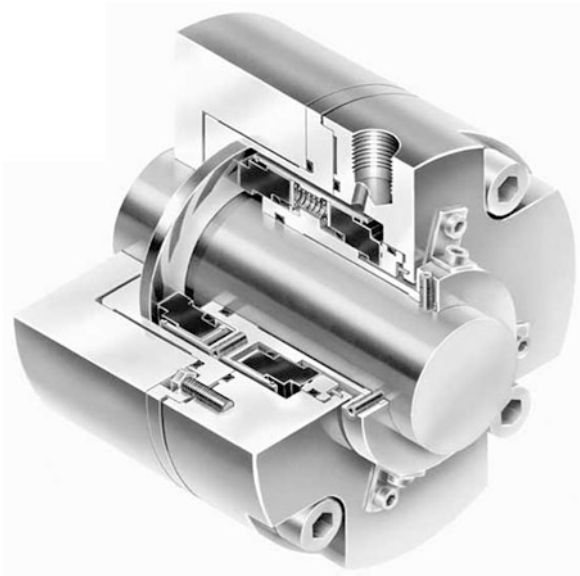
Mechanical Seals, Fig. 12 Type A single seal



Mechanical Seals, Fig. 13 Low-temperature metal bellows seal



Mechanical Seals, Fig. 14 High-temperature metal bellows seal

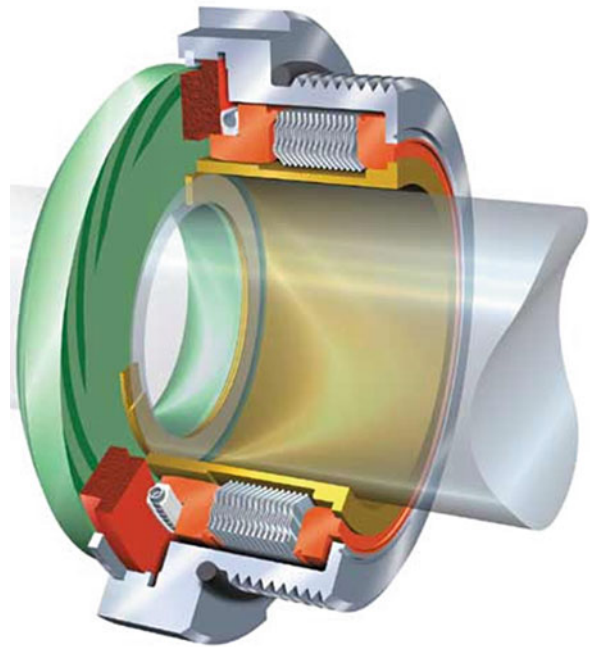


Mechanical Seals, Fig. 15 Gas-lubricated double seal
(John Crane, Inc.)

These applications require a gas-lubricated seal system. This patented seal design is shown in [Fig. 15](#).

Nitrogen or purified air is injected into the seal chamber at higher pressure than the product in the pump. The advantages of this design are:

1. Product in the pump cannot escape to the environment.
2. A pump can run dry without damaging the seals.
3. The gas-lubricated seals protect the product in the pump from liquid contamination.
4. The heat developed from contacting seals is eliminated so the pump can be used to move heat-sensitive fluids.



Mechanical Seals, Fig. 16 Non-contacting seal for cryogenic service (John Crane, Inc.)

This technology has improved equipment reliability and reduced operating cost.

Cryogenic Service

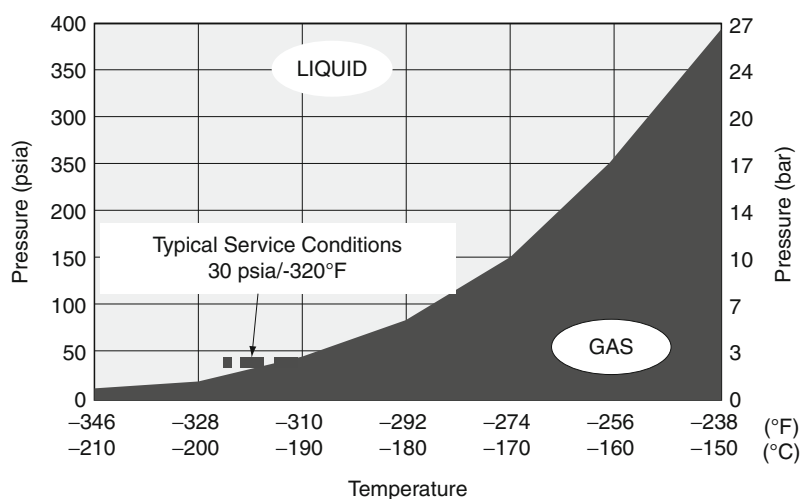
When a seal is required to operate near the boiling point of the liquid sealed, the design must be such that frictional heat is completely eliminated. Any heat developed will cause the liquid to flash directly to a gas. A violent change of state will cause extreme wear and, in some cases, heat checking of the mating ring.

By design, a non-contacting seal has been developed to allow this change of state without damage to the seal. This non-contacting seal for cryogenic service is shown in Fig. 16.

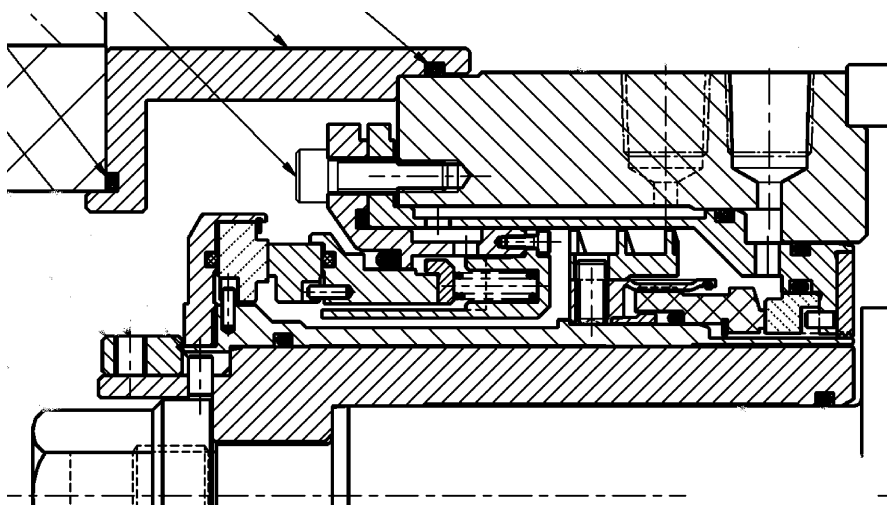
These seals have been applied to pumps on tank trucks handling liquid argon, nitrogen, and oxygen. The temperature range from a solid to a gas for argon, nitrogen, and oxygen is 6, 25, and 65°F, respectively. These are the most difficult fluids to seal with a conventional contacting seal. The success of this sealing concept can be explained by reviewing the vapor pressure curve for nitrogen, Fig. 17.

Nitrogen is normally transported by tanker truck at a pressure of 2 bar (30 psig) and at a temperature of -190°C (-320°F).

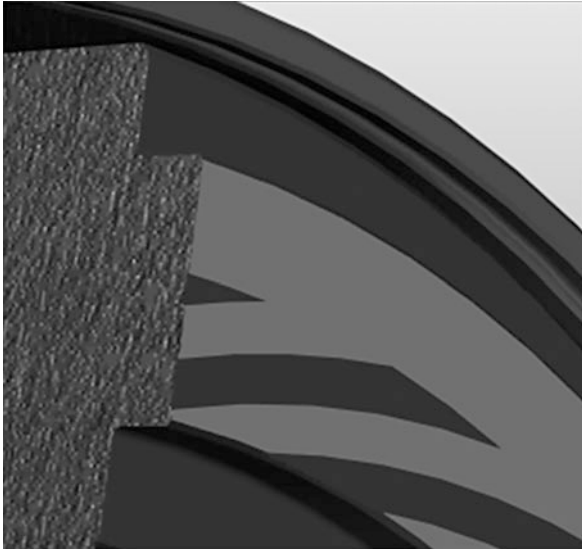
When using a contacting seal, the temperature increase at the seal faces is sufficient to start the boiling process at pumping temperature and pressure. In an uncontrolled environment, damage to the seal faces will occur. Through the proper design of a non-contacting seal with spiral groove seal faces, seal life was increased from 6 weeks to more than 8 years. This resulted in significant savings to the owner of the fleet of trucks.



Mechanical Seals, Fig. 17 Vapor pressure curve for nitrogen



Mechanical Seals, Fig. 18 Spiral groove seal face for active lift (John Crane, Inc.)



Mechanical Seals, Fig. 19 Active lift seal design (John Crane, Inc.)

Active Lift Seals

Active lift or upstream pumping refers to a seal face design that can move a very small quantity of liquid from a low pressure to a high pressure. This, again, is a very significant advancement in seal technology. This design is also based on spiral groove seal technology, [Fig. 18](#).

By controlling the liquid film thickness across the entire face width, the power loss is significantly reduced. The seal faces are not in contact. Conventional pressure–velocity limitations for contacting seals and resultant wear do not apply. This seal must be used in a double seal arrangement, [Fig. 19](#).

The outboard seal contains the liquid in the seal chamber. The inboard seal moves a small quantity of liquid across the inboard seal face, keeping the seal faces clean. On a very difficult service where the liquid being pumped was water, oil, and sand slurry (20% wt.), clean water was moved across the seal faces to a maximum pressure of 40 bar (590 psig). This increase in pressure was achieved with a 75 mm (2.95 in.) seal rotating at 3,600 rpm. Even higher pressures can be achieved by changing the seal face design.

Summary

Every area of our modern society is dependent on rotating equipment. In transportation, mechanical seals are used in automotive, trucking, air travel, and marine craft.

Mechanical seals play a major role in processing industries such as chemicals and refining to provide us with

products that improve the quality of life. The power industry is dependent on seal technology to provide us with our electrical energy needs. Lastly, the food industry, both farm and food processing, depends on seals for the products they produce. Given the wide variety of fluids to be sealed and the infinite number of operating conditions, mechanical seals provide excellent reliability and value to the user.

Cross-References

- [Dry Gas Seals](#)
- [Mechanical Seal Analysis](#)

References

- R.P. Gabriel, S.K. Niamathullah, Design and testing of seals to meeting API 682 requirements, in *Proceedings of the 13th International Pump Users Symposium*, (Turbo-Machinery Laboratory, Texas A&M University, College Station, March 1996)
- A.O. LeBeck, *Principles and Design of Mechanical Face Seals*, (John Wiley and Son, Inc., New York, 1991)
- J.L. Morton, J. Attard, J.G. Evans, Active lift seal technology impact on water injection services, in *22nd International Pump Users Symposium* (Turbo-machinery Laboratory, Texas A & M University, College Station, March 2005)
- J.P. Netzel, Chapter 2.2.3., Centrifugal Pump Mechanical Seals, in *Pump Handbook*, ed. by I.J. Karassik, J.P. Messina, P. Cooper, C.C. Heald, 3rd edn. (McGraw-Hill, New York, 2001), pp. 2.197–2.237
- J.P. Netzel, in High performance gas compressor seals, *11th. International Conference on Fluid Sealing*, BRA, Cranfield, 1987
- J.P. Netzel, J. Voigt, Reducing life-cycle costs for pumps handling cryogenic fluids, in *18th International Pump Users Symposium*. (Turbo-machinery Laboratory, Texas A & M, College Station, 2001)
- K. Schoenherr, *Fundamentals of Mechanical End Face Seals*, in Reprinted from Iron and Steel Engineer, Printed in Engineered Fluid Sealing (John Crane, Inc. 1976)
- K. Schoenherr, R.L. Johnson, Seal Wear, in *Wear Control Handbook*, ed. by M. Peterson, W. Winer (American Society of Mechanical Engineers, New York, 1980)
- API Standard 682, *Shaft Sealing Systems for Centrifugal and Rotary Pumps* (American Petroleum Institute, Washington, DC, 1994)
- J.R. Wasser, R. Sailer, G. Warner, Design and development of gas lubricated seals for pumps, in *11th International Pump Users Symposium*. (Turbo-machinery Laboratory, Texas A&M University College Station, March 1994)

Mechanically Mixed Material

- [Self-Mating Metal Articulations in the Hip Joint](#)

Mechanically Small Cracks

- [Growth Characteristics of Small Fatigue Cracks](#)

Mechanics of Adhesion

- [Contacts Considering Adhesion](#)

Media of Minimum Quantity Lubrication Machining

- [Function of Cutting Fluids and Lubricants](#)

Medium-Range Order (MRO)

- [Crack Growth in Noncrystalline Solids](#)

Medium-Speed Engine Lubricants

- [Locomotive Engine Oils](#)

Medium-Speed Four-Cycle Trunk Piston Engine Oils

- [Marine Engine Oils](#)

Mesh Density Effect on EHL Solution

DONG ZHU

State Key Laboratory of Mechanical Transmission,
Chongqing University, Chongqing, People's Republic
of China

Synonyms

[Grid-convergence of EHL solution](#); [Mesh size effect on EHL solution](#)

Definition

Elastohydrodynamic lubrication (EHL) numerical solution is found to be dependent upon the computational

mesh density used, especially when the film is thin. This article describes the phenomenon, the significance, and practical concerns related to the mesh density effect.

Scientific Fundamentals

Introduction

Elastohydrodynamic lubrication (EHL) is a mode of fluid film lubrication in which hydrodynamic action is enhanced by surface elastic deformation and viscosity increase due to high pressure. This type of lubrication can usually be found in many mechanical components with non-conformal contacts, including various gears, rolling element bearings, cam/follower systems, ball screws, vane pumps, metal rolling tools, and traction drives and continuously variable transmissions. Typically, the EHL contact zone is small, the localized pressure may be as high as 1–4 GPa, the lubricant film very thin, and the contacting surfaces are rough and moving. These create difficulties for experimental studies, therefore, computer simulation is very useful when investigating EHL characteristics. Various numerical solution methods have been developed based on the EHL equation system, which consists of the Reynolds equation for hydrodynamics, an elasticity equation for surface deformation, a geometric equation for contact geometry and film thickness, and lubricant rheology equations for viscosity and density. This equation system is strongly nonlinear in nature, thus, the solution convergence and stability as well as numerical accuracy have been challenging issues to researchers. Refer to “► [EHL Governing Equations](#)” and “► [EHL, Full Numerical Solution Methods](#)” for mathematic details.

In early years, EHL numerical solutions were limited to those in relatively narrow ranges of operating conditions using coarse discretization meshes due to the limited computing power at the time. Because of the convergence problems encountered, the solutions obtained were those usually under light or moderate loads with relatively thick lubricant films (mostly greater than 100–200 nm). The EHL solution did not appear to be very sensitive to the mesh density under such conditions, and computers did not allow very dense meshes in the past, so the coarse meshes were accepted, yielding reasonable results. Based on pioneering studies, such as those by Dowson and Higginson (1959, 1961), for line contact problems, and Hamrock and Dowson (1976, 1977), for point contacts, empirical formulae for predicting film thickness were developed through curve-fitting based on analyzed numerical examples. Those formulae have been widely used in industries since then.

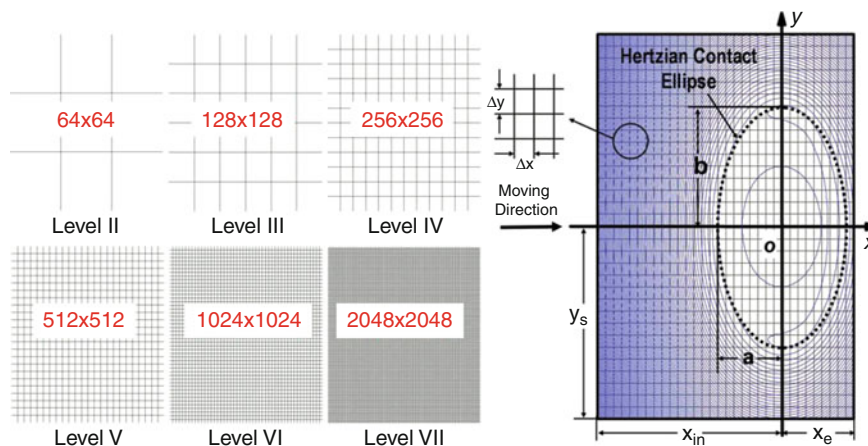
As computer technologies have rapidly advanced, numerical algorithms have improved and mesh density increased. After the mid-1980s, the dimensionless mesh sizes were much reduced from previous levels of 0.06–0.2 down to 0.0075–0.03 by Lubrecht et al. (1987), Venner et al. (1991), Ai et al. (1993), and others. Solution accuracy was improved significantly. The effect of mesh density was studied through numerical examples by different researchers but found to be quite limited.

Since late 1990s, there have been many studies on thin-film and mixed EHL. The thin-film and mixed EHL are of extreme importance not only because most functional components operate in these regimes, but also because they are a critical transition towards lubrication breakdown and surface failures. The mixed EHL numerical solutions have been developed by Zhu and Hu (1999, 2000), Holmes et al. (2005), and others, so that results of thin-film and mixed EHL with asperity contacts can be obtained from a unified numerical approach. Newly developed mixed EHL models are capable of simulating the entire transition from the full-film and mixed EHL down to dry contact under severe conditions. However, previously published mesh effect studies were conducted under thick or moderate film conditions. No mesh effect results for thin films of 20–30 nm or less were found until Morales-Espejel et al. (2005), who studied the mesh effect through their simplified line contact EHL solutions of Grubin's type, and pointed out that the mesh density effect could be insignificant for thick film cases but significant under thin film conditions. Liu et al. 2005, investigated the effects of mesh density and differential schemes on the EHL film thickness through their full numerical solutions in point contacts. Zhu (2007), systematically studied

the mesh effect over a wide range of film thickness, from several hundred nanometers down to a few nanometers and even zero, and further revealed that the film thickness approaches a limit, if mesh density continuously increases, and this asymptotic limit can be readily estimated. The results presented in this essay are mainly from Zhu (2007).

Effect of Mesh Density on Film Thickness

Figure 1 shows a sketch of point contact EHL solution domain in general and sample computational meshes of various density. In order to demonstrate the mesh density effect on the EHL solution, sample cases are given below. Only point contact cases are analyzed as they are more generic, and line contact solutions are simpler, having the same basic trends. The applied load is fixed as 444 N, the material properties are $E' = 226.37$ GPa, $\eta_o = 0.01119$ Pa.s, and $\alpha = 14.94$ GPa⁻¹ (note that a Newtonian fluid is assumed and the Barus exponential pressure-viscosity relation is employed). The geometric parameters are $R_x = R_y = 6.35$ mm, so that the Hertzian contact zone is a circular area of radius $a = 0.2653$ mm, and the maximum Hertzian pressure is $P_h = 3.011$ GPa. The rolling velocity varies from $U = 0.00005$ to 50 m/s, so that the study consequently covers a wide range of central film thickness from zero up to several hundred nanometers. Because most attention is paid to the ultra-thin film region, where lubricant starvation effect caused by small inlet and lateral distances is insignificant and is of no importance to the present study, a small solution domain of $-1.2 \leq x/a \leq 1.1$ and $-1.15 \leq y/a \leq 1.15$ is used in order to achieve a maximal mesh density with certain numbers of grids in both the x- and y-directions.



Mesh Density Effect on EHL Solution, Fig. 1 Solution domain and meshes on different levels of density

Mesh Density Effect on EHL Solution, Table 1 Mesh parameters for sample cases

Mesh level K ($M \times N$)	Mesh size $\Delta X = \Delta Y$	Mesh density D_M
II (64×64)	0.035938	27.826
III (128×128)	0.017969	55.652
IV (256×256)	0.0089844	111.30
V (512×512)	0.0044922	222.61
VI ($1,024 \times 1,024$)	0.0022461	445.22
VII ($2,048 \times 2,048$)	0.0011230	890.44

A direct iterative solution method is used, and in each case the calculation is started from a low grid level, $K = \text{II}$ (mesh numbers: 64×64). After a sufficient number of iterations, the central film thickness, h_c , becomes stabilized. Upon recording converged $H_k = h_c$ for level K mesh, the calculation goes to the next grid level with a doubled mesh density. This procedure is repeated until a stabilized solution on the required highest grid level (e.g., $K = \text{VII}$, $2,048 \times 2,048$) is achieved. See Table 1 for detailed mesh parameters.

Figure 2 summarizes results from a typical sample case, $U = 0.5$ m/s, showing that, generally, the converged central film thickness depends considerably on the mesh used. It is observed that, as the mesh density continuously increases, the converged film thickness is stepwise increased, and the difference, Δ , in converged film thickness between two subsequent mesh levels becomes smaller and smaller. If Δ_K is defined as the difference between the converged film thickness values from mesh levels K and $K+1$, $\Delta_K = H_{K+1} - H_K$, then we have approximately $\Delta_{K+1} = \Delta_K/4$, due to the widely used second-order differential scheme while solving the Reynolds equation. As a result, the film thickness quickly approaches a certain limit when the mesh density is continuously increased. If H_∞ is defined as the converged film thickness obtained from an infinitely dense mesh, it can be estimated as follows:

$$H_\infty = H_K + \Delta_K + \frac{\Delta_K}{4} + \frac{\Delta_K}{16} + \frac{\Delta_K}{64} + \frac{\Delta_K}{256} + \frac{\Delta_K}{1024} + \dots$$

.....
An infinite decreasing geometric series

$$= H_K + \frac{4\Delta_K}{3} = H_{K+1} + \frac{H_{K+1} - H_K}{3}$$

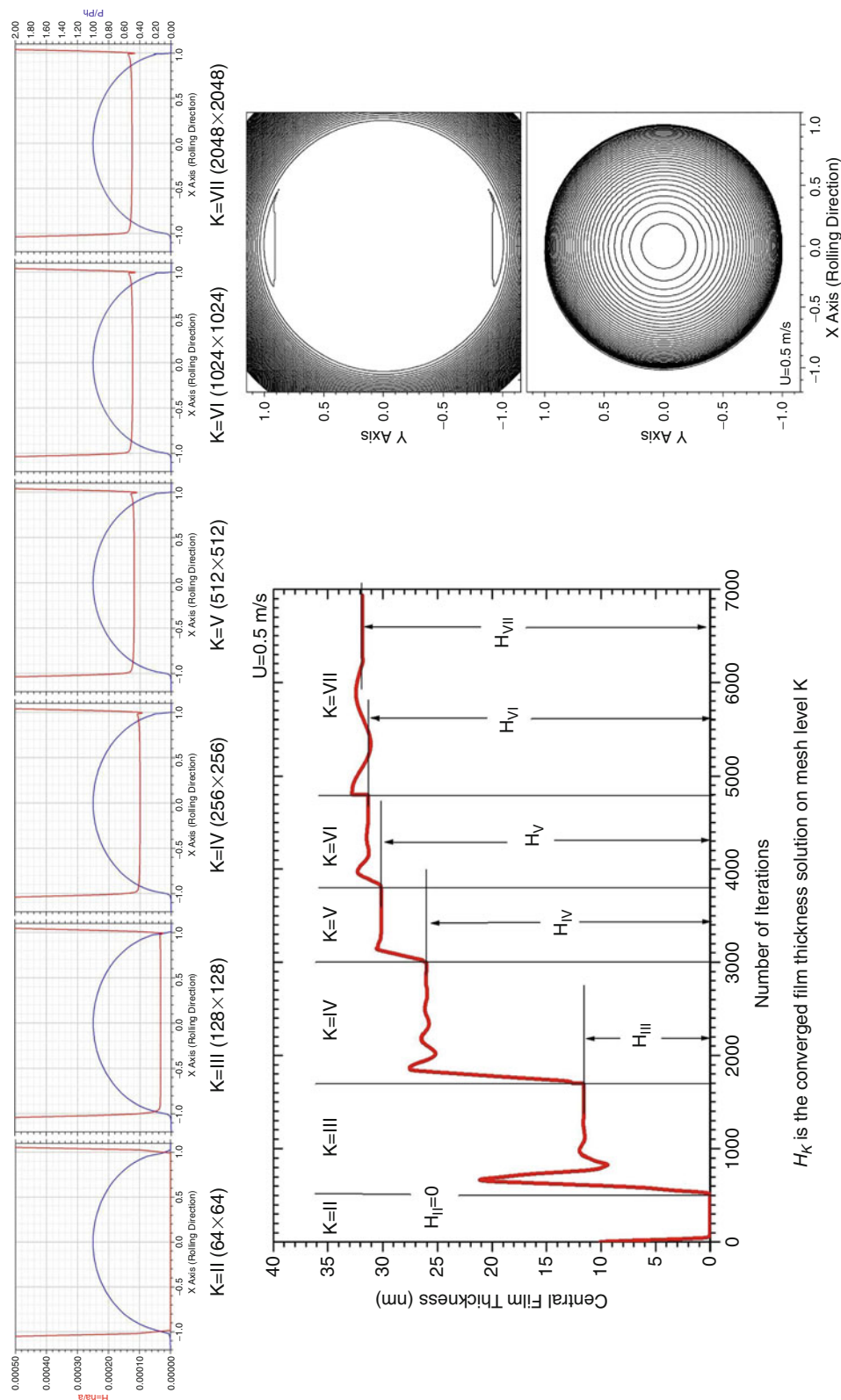
This means that, as soon as two converged non-zero film thickness solutions on subsequent mesh levels are available, one can readily estimate the “theoretically accurate” film thickness, H_∞ , that is supposed to be from the infinitely dense mesh. Note that “theoretically accurate”

means it is no longer dependent on the discretization mesh density used, but numerical truncation errors may still exist. This is of great importance, because, as mentioned above, the accurate film thickness depends on the mesh density, but it is undesirable to use an extremely dense mesh in engineering practice.

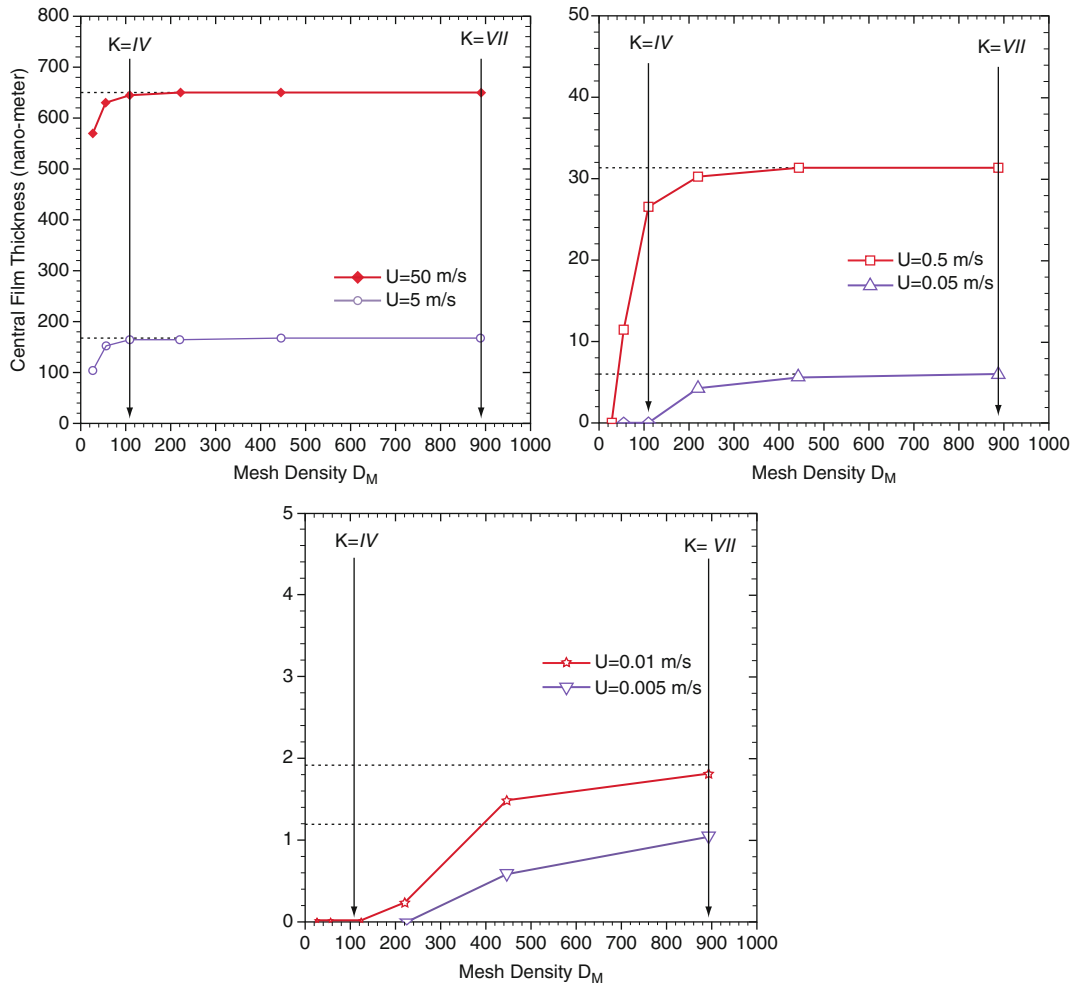
Figure 3 demonstrates the mesh density effect on the central film thickness with obtained results plotted in a different way. Note that for each case the dotted solid line is from the numerical solutions, while the horizontal dashed line illustrates the asymptotic limit, H_∞ . The results show that, for thick film cases, such as those of $U = 50$ m/s and 5 m/s, in which the film thickness is greater than 100 nm, any mesh densification beyond Level V (512×512) only causes a film thickness increase of less than 1 nm (less than 1%). This means that Level V mesh is already sufficiently dense. Even Level IV mesh (256×256) may also be considered sufficient in engineering practice, as the difference between H_{IV} and H_∞ is still within 1–3%. For cases of moderate film thickness, such as that at $U = 0.5$ m/s ($H_\infty = 31.53$ nm), the Level IV mesh yields a relative error in film thickness as great as 15%. However, with the Level V mesh, this relative error is quickly reduced to about 3.9%, and with Level VI less than 1%. When the calculated film thickness becomes tiny, say less than 5 nm in the graph on the right hand side, the relative errors with Level IV and V meshes become mathematically greater. However, the accuracy improvement effort through mesh densification beyond Level V only yields a limited increase of calculated film thickness around or less than 0.5 nm, which is much smaller than the molecule sizes of most industrial lubricants and comparable to (or smaller than) the atomic sizes of elements of interest. Since the calculated tiny film thickness and its limited accuracy improvement based on the continuum mechanics do no longer actually reflect the real interfacial physics on a molecular or atomic scale (or even sub-atomic scale), it does not seem to be necessary, in most cases, to pursue such numerical accuracy improvement with an extremely dense mesh that causes tremendously increased computing time.

Recommended Mesh Density

It is well known that, for any computational analysis based on numerical discretization, the higher the mesh density, the smaller the discretization error but the longer the computing time, if the model and the numerical scheme remain the same. Specifically, for the EHL problems, when mesh density is doubled, the change of converged film thickness solution between two subsequent meshes reduces by a factor of 4, but the computation time



Mesh Density Effect on EHL Solution, Fig. 2 Converged film thickness solution depending on mesh level K



Mesh Density Effect on EHL Solution, Fig. 3 Effect of mesh density on central film thickness. Horizontal dashed lines illustrate estimated asymptotic limits, H_∞

required for each iteration is increased approximately by the same factor of 4, if a second-order discretization scheme is used. Therefore, it is always needed in practice to find a reasonably good compromise between the numerical accuracy and the computation efficiency. The usage of extremely dense meshes should be avoided unless absolutely necessary.

It can be concluded, based on the analysis results and discussion above, that for most engineering research and application projects, a mesh density around $D_M = 80$ – 200 , corresponding to Level IV or V mesh in the above cases, can be used, and the resultant numerical errors are acceptable (say within a few percent). When extra caution is needed, an increased mesh density, $D_M = 250$ – 500 (e.g. Level VI in the above cases), may be more desirable.

It is usually not necessary to employ extremely dense meshes, $D_M > 500$ (beyond Level VI above). In addition, a small solution domain is often recommended in order to maximize the mesh density, if the study is focused on the thin film region where starvation effect is insignificant. Recommendations given above are mainly for point contact problems. The mesh can be further increased for line contact problems, as they require much less computing time.

Application

EHL film thickness is an important parameter representing lubrication effectiveness in mechanical components. Prediction of EHL film thickness, therefore, is vital to performance and efficiency improvements and

failure prevention through design optimization. It has been found, however, that the film thickness predicted through numerical solution is dependent upon the computational mesh used, especially when the film is thin. In engineering practice, it is impossible, or at least undesirable, to use extremely dense meshes, which require huge computing time and memory space. It has been found, fortunately, that the converged film thickness solution approaches an asymptotic limit as the mesh density continuously increases. This limit, H_∞ , is no longer dependent on the mesh density, and it can be readily estimated by using the following relationship, if the second order discretization scheme is used when solving the EHL problems:

$$H_\infty = H_{k+1} + \frac{H_{k+1} - H_k}{3}$$

This means that the theoretically accurate film thickness, H_∞ , can be predicted with two converged non-zero film thickness solutions from subsequent meshes, which do not have to be highly dense. This could save considerable computing time in practice.

Also, when solving the EHL equation system numerically, one should carefully choose the mesh density, having a good compromise between the numerical accuracy and the computation efficiency. Generally a mesh density of around 80 ~ 200 can be considered sufficient for engineering analyses. When extra caution is needed, a high density of 250 ~ 500 can be used. Mesh density greater than 500 (which may require huge computing time and memory space) should be avoided, unless absolutely necessary under unusual circumstances. The above recommended mesh densities are mainly for point contact problems. The mesh can be further densified when solving line contact problems, which require much less computing time.

Nomenclature

a	Semi-axis of Hertzian ellipse in x-direction, or radius of Hertzian circle
D_M	Mesh density parameter, defined as the total number of meshes within the distance of Hertzian radius, a
E'	Effective elastic modulus
h	Local film thickness (or gap)
h_c	Central film thickness
$H_I, H_{II}, H_{III}, H_{IV}, \dots$	Converged central film thickness solutions on mesh levels I, II, III, and IV. . . , respectively
H_∞	Central film thickness theoretically from an infinitely dense mesh

K	Mesh level
p	Pressure
P_h	Maximum Hertzian pressure
R_x, R_y	Effective radii of curvature in x- and y- directions, respectively
U	Rolling velocity (or entraining velocity)
x, y	Coordinates (x is chosen to be parallel to rolling direction)
$X, Y = x/a, Y = y/a$	Dimensionless coordinates
α	Pressure-viscosity exponent used in viscosity equation $\eta = \eta_0 e^{\alpha p}$
$\Delta x, \Delta y$	Mesh sizes in x- and y-directions, respectively
$\Delta X, \Delta Y = \Delta x/a, \Delta Y = \Delta y/a$	Dimensionless mesh sizes
η, η_0	Viscosity and viscosity under ambient condition, respectively

Cross-References

- [EHL, Full Numerical Solution Methods](#)
- [Elastohydrodynamic Lubrication \(EHL\)](#)
- [Multi-Grid Method](#)
- [Progressive Mesh Densification \(PMD\) Method](#)

References

- X. Ai, Numerical analyses of elastohydrodynamically lubricated line and point contacts with rough surfaces by using semi-system and multigrid methods. Ph.D. thesis, Northwestern University, Evanston, 1993
- D. Dowson, G.R. Higginson, A numerical solution to the elastohydrodynamic problem. *J. Eng. Sci.* **1**, 6–15 (1959)
- D. Dowson, G.R. Higginson, New roller bearing lubrication formula. *Engineering* **192**, 158–159 (1961)
- B.J. Hamrock, D. Dowson, Isothermal elastohydrodynamic lubrication of point contacts, part 1 – theoretical formulation. *ASME J. Lubr. Technol.* **98**, 223–229 (1976)
- B.J. Hamrock, D. Dowson, Isothermal elastohydrodynamic lubrication of point contacts, part 3 – fully flooded results. *J. Lubr. Technol.* **99**, 264–276 (1977)
- M.J.A. Holmes, H.P. Evans, R.W. Snidle, Analysis of mixed lubrication effects in simulated gear tooth contacts. *J. Tribol.* **127**, 61–69 (2005)
- Y.Z. Hu, D. Zhu, A full numerical solution to the mixed lubrication in point contacts. *J. Tribol.* **122**, 1–9 (2000)
- Y. Liu, Q. Wang, W.Z. Wang, Y.Z. Hu, D. Zhu, Effects of differential scheme and mesh density on EHL film thickness in point contacts. *J. Tribol.* **128**, 641–653 (2005)
- A. A. Lubrecht, The numerical solution of elastohydrodynamic lubricated line and point contact problems using multigrid techniques. Ph.D. Thesis, University of Twente, The Netherlands, 1987
- G.E. Morales-Espejel, M.L. Dumont, P.M. Lugt, A.V. Olver, A limiting solution for the dependence of film thickness on velocity in EHL contacts with very thin films. *Tribol. Trans.* **48**, 317–324 (2005)

- C. H. Venner, Multilevel solution of EHL line and point contact problems. Ph.D Thesis, University of Twente, The Netherlands, 1991
- D. Zhu, On some aspects in numerical solution of thin-film and mixed EHL. *Proc. IMech, Part J, J.Eng. Tribol.* **221**, 561–579 (2007)
- D. Zhu, Y.Z. Hu, The study of transition from full film elastohydrodynamic to mixed and boundary lubrication, in *The Advancing Frontier of Engineering Tribology, Proceedings of the 1999 STLE/ASME H.S. Cheng Tribology Surveillance*, Park Ridge, Illinois, USA, pp. 150–156, 1999

Mesh Size Effect on EHL Solution

► Mesh Density Effect on EHL Solution

Meshfree Method for Contact Simulation

GENG LIU

Mechanical Engineering, Northwestern Polytechnical University, Xian, Shaanxi, People's Republic of China

Synonyms

Meshless or element-free method for contact analysis

Definition

Meshfree method for contact simulation is an analysis procedure of contact physical phenomena by using a particular class of numerical simulation algorithms that establish system algebraic equations for the whole calculation domain without using a predefined grid or mesh for the domain discretization.

Scientific Fundamentals

Contact Problems and Meshfree Method

Contact problems are usually nonlinear, either in terms of geometry or elasticity, or both (Johnson 1985). Solving contact problems with rough surfaces is in principle of the greatest challenge. Although most available methods for the contact analysis are by means of theoretic approaches, either using a Hertzian type of formulas or integral solutions by means of Greens functions, the finite element method has been used to solve the contact problems of complex shapes and even with rough surfaces. However, when considering the large deformation in the real area of contact of asperities, the meshes for the finite elements

may be twisted. The use of the finite element method to solve contact problems faces a dilemma: how the surface distortion should be properly taken into account.

The meshfree method seems to be a promising approach for contact analyses because of its flexibility in domain discretization and versatility in node arrangements. It has been used to solve a variety of complicated engineering problems, such as cracks, shocks, and large deformations. The fundamentals of meshfree methods is to replace the meshes employed in the finite element method by an approximation constructed with weight functions for every node in the area analyzed (Belytschko et al. 1996). The mechanical characteristics of a node are influenced by that of the other points in the calculation domain through weight functions.

The meshfree method may be formulated with different algorithms of different weight functions. These algorithms include the smooth particle hydrodynamic (SPH) Method (Lucy 1977), the element-free Galerkin (EFG) method (Belytschko et al. 1994), which is based on the moving least-square (MLS) approximation, the element-free Galerkin-finite element (EFG-FE) coupling method (Belytschko et al. 1995), the partition of unity (PU) method (Babuška and Melenk 1996), the reproducing kernel particle method (RKPM) (W.K. Liu et al. 1995), and the multi-scale method (W.K. Liu et al. 2000), etc.

Solution Equations of Contact Problem

Figure 1 illustrates the elastic contact between an elastic body, Ω_1 , and a rigid plane, Ω_2 , under an applied normal load without considering friction. The contact situation may be expressed by the following displacement and equilibrium conditions.

The displacement conditions

$$\begin{cases} u(\mathbf{x}) + s_0(\mathbf{x}) - \alpha = 0 & \text{within the contact area} \\ u(\mathbf{x}) + s_0(\mathbf{x}) - \alpha > 0 & \text{outside the contact area} \end{cases} \quad (1)$$

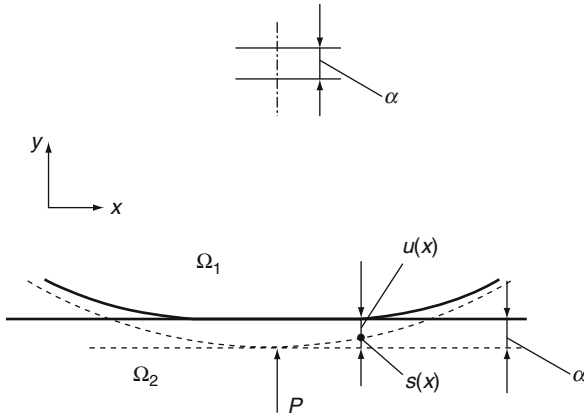
where $u(\mathbf{x})$ is the deformation of the elastic body, $s_0(\mathbf{x})$ the initial surface separation, α the rigid-body displacement.

The equilibrium condition

$$\int_A r(\mathbf{x}) dA = P \quad (2)$$

where $r(\mathbf{x})$ is the contact loads of the contacting bodies, A the area of contact, and P the total applied load.

Linear programming equations can be derived by means of the approximations of the meshfree method, and the simplex-type technique is also used to solve the contact problems. A number of nodes are set in the domain that is potential to be in the contact region. The discrete form of Eqs. 1 and 2 can be rewritten as



Meshfree Method for Contact Simulation, Fig. 1 The deformed relationships of two contact bodies

The displacement condition

$$-\mathbf{u} + \alpha \mathbf{e} + \mathbf{y} \mathbf{I} = \varepsilon_0 \quad (3)$$

The equilibrium condition

$$\mathbf{r}^T \mathbf{e} = P \quad (4)$$

where \mathbf{u} is the deformation of possible contact nodes, \mathbf{e} an unit vector, \mathbf{y} the slack vector, \mathbf{I} the identity matrix, ε_0 the initial surface separation vector of possible contact nodes, and \mathbf{r} the nodal contact loads in the contacting bodies. The displacement, \mathbf{u} , is computed by

$$\mathbf{u} = \mathbf{F} \mathbf{r} \quad (5)$$

where \mathbf{F} is the flexibility matrix, or the influence function matrix, for possible contact nodes in the two contacting bodies.

A slack vector, \mathbf{y} , is introduced to (Eq. 3) to simplify the contact criterion

$$\begin{cases} \text{If } y_i = 0, & r_i \geq 0 & \text{within the contact area.} \\ \text{If } y_i > 0, & r_i = 0 & \text{outside the contact area.} \end{cases} \quad (6)$$

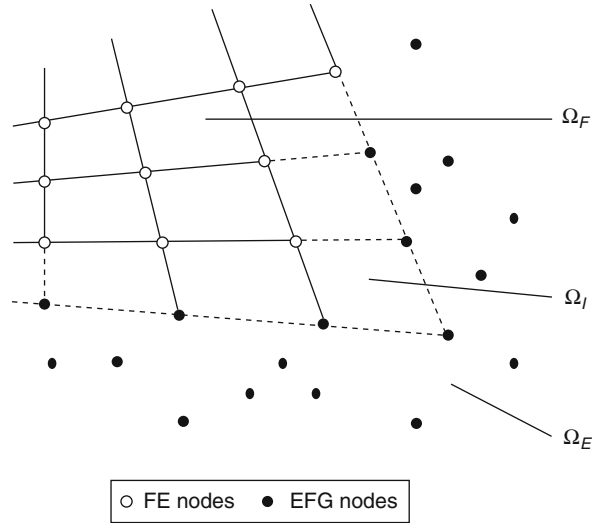
$(i = 1, 2, \dots, n)$

where n is the number of nodes in the possible region of the surface in contact.

The solution to the contact problem is to solve the contact load distribution, \mathbf{r} , for contact nodes from Eqs. 3, 4, and 6, and then to calculate the mechanical performance of contact bodies with respect to the contact load.

Approximations of the EFG-FE Method

The meshfree method requires more computer resources than the conventional finite element method does because its weight function involves more nodes in the influenced



Meshfree Method for Contact Simulation, Fig. 2 Element-free Galerkin-finite element coupling method

domain. In order to save CPU time, the meshfree area of a model must be controlled as small as possible. An element-free Galerkin-finite element coupling method has been developed by Belytschko (Belytschko et al. 1995). Using this coupling method, the meshfree method is applied only in the regions where the finite element method is too coarse to be used.

Interface elements shown in Fig. 2 (the dash line elements) are introduced to couple the EFG nodes and the finite element mesh and ensure continuity in the computation. In Fig. 2, Ω_E expresses the EFG domain, Ω_I the interface domain and Ω_F the FE domain.

By setting a number of nodes in a domain, as shown in Fig. 2, the displacement approximation can be defined as

$$u^h(\mathbf{x}) = \sum_I \bar{N}_I(\mathbf{x}) \mathbf{u}_I \quad (7)$$

where \mathbf{u}_I is the displacement vector of the nodes, $\bar{N}_I(\mathbf{x})$ the shape function of the coupling method which can be obtained as follow

$$\bar{N}_I(\mathbf{x}) = \begin{cases} N_I(\mathbf{x}) & \mathbf{x} \in \Omega_I \\ \Phi_I(\mathbf{x}) & \mathbf{x} \in \Omega_E \\ \tilde{N}_I(\mathbf{x}) & \mathbf{x} \in \Omega_F \end{cases} \quad (8)$$

where $N_I(\mathbf{x})$ is the FE shape function, $\Phi_I(\mathbf{x})$ the EFG shape function, $\tilde{N}_I(\mathbf{x})$ the shape function of interface elements. And the EFG shape function is

$$\Phi^k = \mathbf{p}^T(\mathbf{x}) \mathbf{A}^{-1}(\mathbf{x}) \mathbf{B}(\mathbf{x}) \quad (9)$$

where $\mathbf{A} = \mathbf{P}^T \mathbf{W}(\mathbf{x}) \mathbf{P}$, $\mathbf{B} = \mathbf{P}^T \mathbf{W}(\mathbf{x})$, \mathbf{P} is a base matrix formed by the base functions at all nodes in the domain and \mathbf{W} is the weight matrix

$$\mathbf{P} = \begin{bmatrix} p_1(\mathbf{x}_1) & p_2(\mathbf{x}_1) & \dots & p_m(\mathbf{x}_1) \\ p_1(\mathbf{x}_2) & p_2(\mathbf{x}_2) & \dots & p_m(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(\mathbf{x}_n) & p_2(\mathbf{x}_n) & \dots & p_m(\mathbf{x}_n) \end{bmatrix},$$

$$\mathbf{W} = \begin{bmatrix} w(\mathbf{x} - \mathbf{x}_1) & 0 & \dots & 0 \\ 0 & w(\mathbf{x} - \mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w(\mathbf{x} - \mathbf{x}_n) \end{bmatrix}$$

in which $w(\mathbf{x} - \mathbf{x}_I)$ is a weight function and is defined by Monaghan (Monaghan, 1982).

The partial derivatives of Φ^k can be obtained as

$$\Phi_{,i}^k = \mathbf{p}_{,i}^T \mathbf{A}^{-1} \mathbf{B} + \mathbf{p}^T (\mathbf{A}_{,i}^{-1} \mathbf{B} + \mathbf{A}^{-1} \mathbf{B}_{,i}) \quad (10)$$

where $\mathbf{A}_{,i}^{-1} = -\mathbf{A}^{-1} \mathbf{A}_{,i} \mathbf{A}^{-1}$.

The shape function, Φ^k , depends on the base and weight function. It plays a key role in the numerical operation of the element-free Galerkin method. The selection of the weight function is crucial to achieve the best numerical results.

The interface shape function is

$$\tilde{N}_I(\mathbf{x}) = [1 - R(\mathbf{x})]N_I(\mathbf{x}) + R(\mathbf{x})\Phi_I(\mathbf{x}) \quad (11)$$

The partial derivatives of $\tilde{N}_I(\mathbf{x})$ can be obtained as

$$\tilde{N}_{I,i} = -R_{,i}N_I + (1 - R)N_{I,i} + R_{,i}\Phi_I + R\Phi_{I,i} \quad (12)$$

where $R(\mathbf{x}) = \sum_{\mathbf{x}_j \in \Gamma_{EFG}} N_j(\mathbf{x})$, $R_{,i} = \sum_{\mathbf{x}_j \in \Gamma_{EFG}} N_{j,i}$ and

N_j is the Lagrange interpolate.

The discrete equation of the coupling method is

$$\mathbf{K}\mathbf{u} = \mathbf{f} \quad (13)$$

where

$$\mathbf{K}_{IJ} = \int_{\Omega} \mathbf{B}_I^T \mathbf{D} \mathbf{B}_J d\Omega \quad (14)$$

$$\mathbf{f} = \int_{\Gamma_t} \bar{N}_I \bar{\mathbf{t}} d\Gamma + \int_{\Omega} \bar{N}_I \mathbf{b} d\Omega \quad (15)$$

in Eq. 14

$$\mathbf{B}_I = \begin{bmatrix} \bar{N}_{I,x} & 0 \\ 0 & \bar{N}_{I,y} \\ \bar{N}_{I,y} & \bar{N}_{I,x} \end{bmatrix} \quad (16)$$

Weight Function and Numerical Integration

The meshfree method is not based on element structures usually used in the finite element method but the support of the weight function. The weights of the meshfree method are usually functions of distance. If the distance is $d_I = \|\mathbf{x} - \mathbf{x}_I\|$, then

$$w(\mathbf{x} - \mathbf{x}_I) = w(d_I) \quad (17)$$

According to Monaghan (Monaghan, 1982), there are many choices for the weight function, such as exponential, cubic, or quadric spline function. The exponential weight function is often used,

$$w(d_I) = \begin{cases} \frac{e^{-(d_I/c)^2} - e^{-(d_m/c)^2}}{1 - e^{-(d_m/c)^2}} & d_I \leq d_m \\ 0 & d_I > d_m \end{cases} \quad (18)$$

where d_m is the size of the support of the weight function and c is a constant that controls the relative weights, which can be defined as

$$c = \alpha c_I \quad (19)$$

for $1 \leq \alpha \leq 2$ and

$$c_I = \max_{J \in S_j} \|\mathbf{x}_J - \mathbf{x}_I\| \quad (20)$$

where S_j is the minimum set of neighboring points of \mathbf{x}_I , which constructs a polygon surrounding point \mathbf{x}_I . If the nodes in the model are uniformly distributed, c_I in Eq. 19 is the maximum distance between nodes. If the nodes are randomly distributed, c_I can be defined as a characteristic length of the integration zone that contains the point, \mathbf{x}_I . Other choices of c_I are also possible. d_m determines the domain of the influence of \mathbf{x}_I , which may be defined as

$$d_m = \beta c \quad (21)$$

where β is a scaling factor, which is chosen to be $\beta > 2$.

Because the meshfree method does not use an element structure, background cells (Belytschko et al. 1994), as illustrated in Fig. 3, that are independent of the nodes are employed to integrate Eqs. 14 and 15.

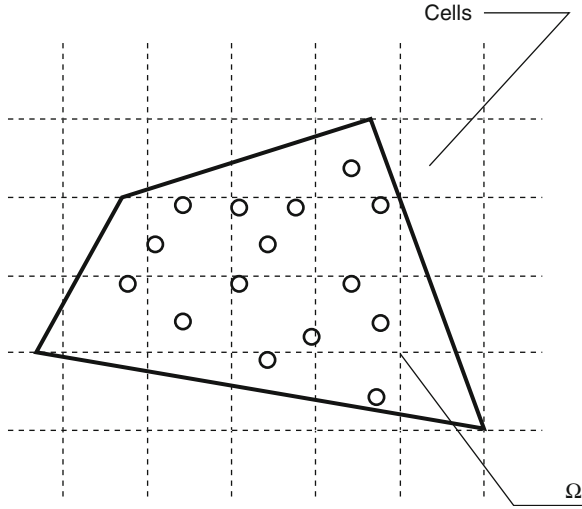
According to Belytschko et al. (1994), the number of the background cells, $m_c \times n_c$, can be determined by

$$m_c = \sqrt{n_N} \quad (22)$$

where n_N is the total number of the nodes. The number of Gauss integral points, $n_Q \times n_Q$, can be determined by,

$$n_Q = \sqrt{m} + 2 \quad (23)$$

where m is the number of the nodes in a cell. Belytschko indicated that it is better to use a lower order of integral in many smaller cells than to use higher order integral in large cells.



Meshfree Method for Contact Simulation, Fig. 3
Background cell Gauss quadrature

Calculation of the Meshfree Node Stresses

In EFG method, if the background cell Gauss integral is used, the stresses of Gauss points can be obtained as

$$\sigma = \mathbf{D} \cdot \mathbf{B} \cdot \mathbf{u} \quad (24)$$

The calculation of the stresses of the meshfree nodes is different from that of the finite element method. An average method based on the support domain of the weight function on Gauss points is used as

$$\sigma_i = \frac{\sum_{j=1}^{n_G} \sigma_j^G}{n_G} \quad (25)$$

where σ_i is the stress vector of the meshfree nodes, σ_j^G is the stress vector of the Gauss points, n_G is the number of the support domain of Gauss points in which the meshfree nodes are included.

Example. The cylinder in contact with a rigid plane is presented to verify the meshfree contact model. The following assumptions are made: (1) the materials of the contact bodies are isotropic, (2) the deformations are elastic, (3) the contact areas are small and the contact surfaces are smooth, (4) the bodies are subjected to a normal load, and (5) there is no friction between the two bodies. These assumptions are the same as those used in the Hertz contact solution. The geometry and material parameters of the cylinder are: radius of the cylinder, $R = 0.5\text{mm}$, Young's modulus, $E = 210\text{GPa}$, Poisson's

ratio, $\nu = 0.3$, and applied load, $P = 50\text{N}$. The contact problem is treated as a plane-strain problem.

Due to the geometric symmetry, one half of the cylindrical structure is established as shown in Fig. 4, where the meshfree region is discretized into 1,609 nodes. The space integrations are performed with the assistance of 80×40 background cells in model. In the calculation, the support of the weight function, $d_m = 5c$, and the number of Gauss integral points is 2×2 .

The relationships between the nondimensional contact pressure, P/P_0 , and the nondimensional half width, x/a , are shown in Fig. 5. The theoretical solution to such an elastic cylinder-plane contact problem indicates that the maximum contact pressure is $P_0 = 2710\text{MPa}$ and that the half width of the contact is $a = 0.0117\text{mm}$, respectively. The solution methods are validated by excellent agreement between the theoretical and numerical results for the contact pressures and areas.

Key Applications

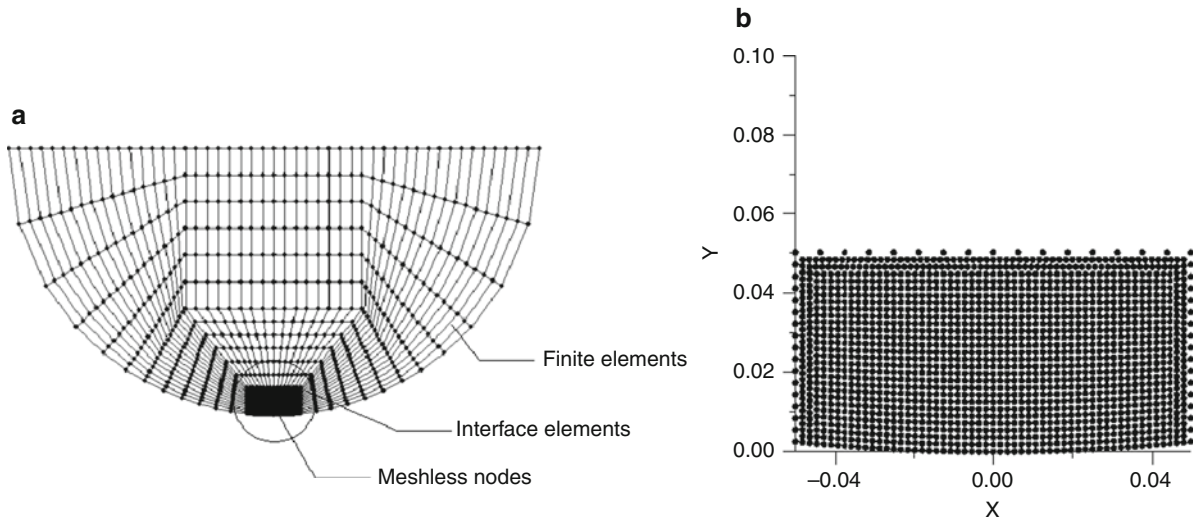
Micro Scale Adhesive Contact Problems

Adhesion may occur when bodies of small scales approach to each other. Understanding the properties of adhesive contact requires mathematic descriptions of the geometrical characteristics of the surfaces of two bodies in contact. The surface effect and the scale effect are the two major concerns among several basic problems to be solved in the study of micro scale adhesive contact. Some of the factors, such as the van der Waals force, inter-atomic force, and electrostatic force, etc., which usually ignored in conventional macro contact analyses, may become noticeable when the length scale is reduced.

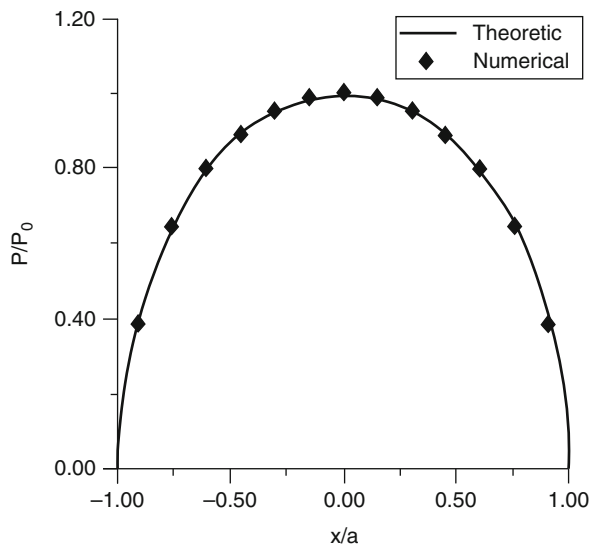
The meshfree method can be used to solve the adhesive contact with more complex surface topography and to consider more intricate factors, such as thermal stress, friction, elasto-plastic deformation, etc.

Adaptive Meshfree Method for Elasto-Plastic Contacts

For many common situations of contact problems, such as stress concentration and single-point contact, some physical quantities change dramatically in a very small region along with the great variations in local area or boundary position. This kind of problem causes great difficulties and challenges to numerical calculation. If uniform meshes are employed to solve this kind of problem, then it is usually unlikely to acquire a result with reasonable accuracy and it sometimes even makes the amount of computation beyond the computer capacity. The adaptive method



Meshfree Method for Contact Simulation, Fig. 4 The arrangement of nodes placing in half cylinder. (a) Structure of EFG-FE coupling. (b) Nodes placing in meshfree region



Meshfree Method for Contact Simulation, Fig. 5 The contact pressures of theoretical and numerical solution

could effectively eliminate numerical result oscillation caused by acute change of local physical quantities and consequently reduce computing errors, improve numerical stability, and finally enhance the economy of computation on the premises of ensuring reasonable accuracy. Considering the meshfree method is not confined to FE meshes and their discretization, it is particularly

suitable for adaptive analysis because of its convenience in adding or removing nodes in the calculating region. Using meshfree method for adaptive analysis, a more accurate numerical model is readily obtained by just inserting discrete nodes in the high strain gradient field. Zhang et al. indicated that the accuracy of the solutions from the meshfree adaptively refined is satisfactory, while that the CPU time cost is only about 20% of the uniformly refined calculation (Zhang et al. 2008).

Adaptive-Surface-Based Asperity Contact

Real surfaces are not perfectly smooth which consist of many micro irregular protuberances. The description of rough surface is one of the key factors which may influence the performance analyses for asperity contacts. In general, the surface topography can be characterized by experimental measurement, such as using the stylus-type surface profilometers, scanning electron microscope (SEM), scanning tunneling microscope (STM), etc. It could also be characterized through mathematical description, such as GW statistical contact model, fractal model, etc. Mostly, a 2D surface profile is described by a series of discrete nodes with the same intervals along the coordinate axis no matter the surface is generated by mathematical methods or is digitized through measuring instruments.

In computer graphics, much research has concerned the adaptive description of surface topography that reduces the number of controlling nodes on surface

while maintaining the accuracy of interpolation. Based on these ideas, the rough surfaces in contact can also be adaptively described through which the nodes are dense on the rough parts of the surfaces but coarse on the smooth parts. So, the computing time can be greatly reduced by removing the nodes on the surface that have little influence on the contact behaviors of rough surfaces. The removed nodes are determined by setting a threshold. Thus, the contact problems can be described by fewer surface nodes but have acceptable results comparing to the ones of the original surface. However, if the adaptive technique is used, it is somehow difficult to construct the FE meshes to adapt the surface description. The adaptive-surface-based asperity contact model can be established perfectly by using the meshfree method because of its flexibility in domain discretization and versatility in node arrangements. Liu et al. indicated that when the relative errors of contact pressures are fewer than 5%, the calculation time of the adaptive-surface-based model is about 50% of the non-adaptive one (T. Liu et al. 2006).

Cross-References

- [Adhesive Contact of Elastic Bodies](#)
- [Contact of Layered Materials](#)
- [History of Contact Mechanics](#)
- [Numerical Methods for Elastic Contact Problems](#)
- [Thermo-Elasto-Plastic Contact](#)

References

- I. Babuška, J.M. Melenk, The partition of unity finite element methods: basic theory and applications. *Comput. Methods Appl. Mech. Eng.* **139**, 91 (1996)
- T. Belytschko, Y.Y. Lu, L. Gu, Element-free Galerkin methods. *Int. J. Numer. Methods Eng.* **37**, 229 (1994)
- T. Belytschko, D. Organ, Y. Krongauz, A coupled finite element-element-free Galerkin method. *Comput. Mech.* **17**, 186 (1995)
- T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, P. Krysl, Meshless methods: an overview and recent developments. *Comput. Methods Appl. Mech. Eng.* **139**, 3 (1996)
- K.L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge/New York, 1985)
- W.K. Liu, S. Jun, Y.F. Zhang, Reproducing kernel particle methods. *Int. J. Numer. Methods Eng.* **20**, 1081 (1995)
- W.K. Liu, S. Hao, T. Belytschko et al., Multi-scale methods. *Int. J. Numer. Methods Eng.* **47**, 1343 (2000)
- T. Liu, G. Liu, Q. Xie, Q.J. Wang, Two-dimensional adaptive-surface elasto-plastic asperity contact model. *J. Tribol.* **128**(4), 898 (2006)
- L.B. Lucy, A numerical approach to the testing of the fission hypothesis. *Astron. J.* **8**(12), 1013 (1977)
- J.J. Monaghan, Why particle methods work. *SIAM J. Sci. Stat. Comput.* **3**(4), 422 (1982)
- Z. Zhang, G. Liu, T. Liu, Z. Quan, L. Wu, An adaptive meshless computational system for elastoplastic contact problems. *Int. J. Comput. Methods* **5**(3), 433 (2008)

Meshless or Element-Free Method for Contact Analysis

- [Meshfree Method for Contact Simulation](#)

Mesothelial Tribology

- [Pleural Lubrication and Friction in the Chest](#)

Metal Fiber Brushes

KOICHIRO SAWA

Department of System Design Engineering, Keio University, Kohoku-ku, Yokohama, Japan

Definition

A fiber brush is a specially designed brush for sliding electrical contacts. Generally, fiber brushes are simply groups of individual metal fibers (thin wires) that are collimated by and terminated into a metal tube. The free, unterminated end of the fiber brush bundle rides along a groove on the ring surface.

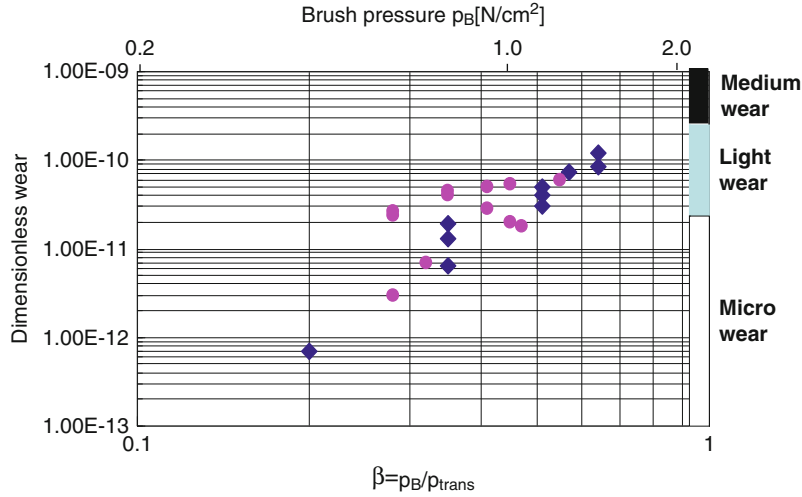
Scientific Fundamentals

The wear explained by the Holm-Archard wear law is an adhesive wear, which is caused by plastic deformation (Archard 1953; Holm 1967). Therefore, a wear rate could be expected to be very low, if sliding contacts are operated under elastic deformation condition.

Figure 1 shows the relation between wear and brush pressure. The dimensionless wear rate is a wear thickness normalized by sliding length. The transition brush load P_{trans} is a critical one between plastic and elastic deformation and obtained by equating the average pressure of the Hertzian contact area A_{el} to the hardness H ;

$$P_{trans} = HA_{el} = 9 \times 10^{-4} H n r_c^2, \quad (1)$$

where n is the number of contact spots and r_c is the average radius of contact asperities.



Metal Fiber Brushes, Fig. 1 Dimensionless wear rate measured for metal fiber brushes of overall 12 mm diameter, constructed of 50 μm diameter copper (●) and silver-clad copper wires (◆) at $f \approx 15\%$ packing fraction, sliding on a polished copper slip ring mostly at 14 ms^{-1} at a 50 Acm^{-2} current density in moist CO_2 (Kuhlmann-Wilsdorf 1996)

Further, the transition pressure p_{trans} is calculated as follows:

$$p_{\text{trans}} = \frac{P_{\text{trans}}}{\text{Apparant contact area}} = 9 \times 10^{-4} H n^* r_c^2, \quad (2)$$

where n^* is the number of contact spots per unit area.

In Fig. 1, β is the ratio of the brush pressure to the critical pressure, p_B/p_{trans} and β_{safe} is defined as $\beta = 1/2$, where p_b is a macroscopic brush pressure, brush load/apparant contact area. From this figure, it is made clear that a wear rate stays in a very low range, micro wear under the pressure below β_{safe} .

In case of a usual monolithic brush, the number of load bearing area more than one and a plastic deformation takes place at the contact spots. A fiber brush is one method to increase the number of contact spots. Figure 2 shows one example of fiber brushes.

A theoretical explanation of fiber brushes is given by Prof. Kuhlmann-Wilsdorf as follows (Kuhlmann-Wilsdorf 1983, 1996; Slade 1999). Two assumptions for the theory are provided: The majority of fibers have one contact per each fiber on the rotor. This is usually realized when the diameter of a fiber is less than 50 μm and the packing ratio of fibers is so low (up to 20%) as not to disturb the deformation of each fiber.

Further, the radius of contact spots is close to that of the fiber, because Fig. 1 indicates that from the value p_{trans} the radius r_c of the asperity forming an elastic, a “Hertzian” contact spot is found to be almost equal to

that of the fiber. Therefore, the approximation leads to the following equation.

$$n^* = f / \left(\frac{1}{4} \pi d^2 \right), \quad (3)$$

where f is packing fraction, percentage of metal in brush volume at interface and d is diameter of the fiber.

Eventually, $r_c = d/2$ and safe pressure p_{safe} ($= p_{\text{trans}}/2$) are introduced and (4) is obtained independent of fiber diameter.

$$p_{\text{trans}} = 2p_{\text{safe}} \approx 3 \times 10^{-4} f H \approx 1.2 \times 10^{-6} f E, \quad (4)$$

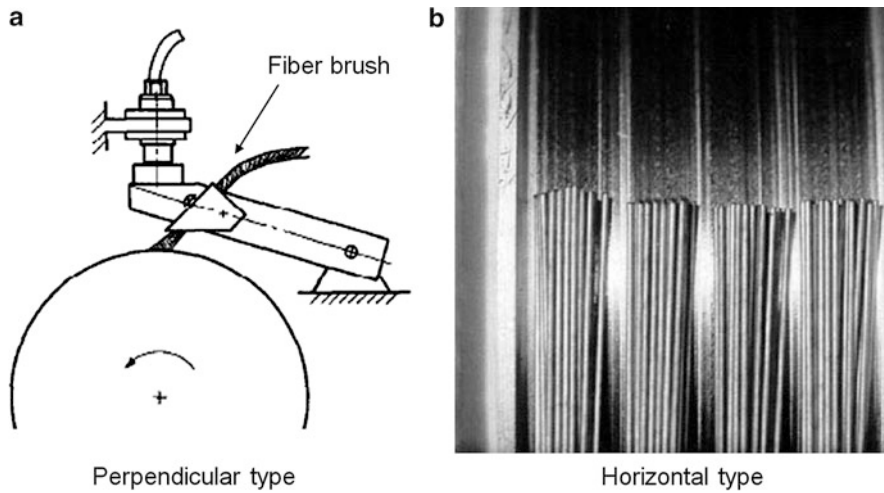
where $H \approx 0.004E$.

The following equation is derived for copper brushes with $E \approx 1.1 \times 10^{11} \text{ N/m}^2$:

$$C_{\text{Cu}} p_{\text{trans}} = 2C_{\text{Cu}} p_{\text{safe}} \approx 15f (\text{Ncm}^{-2}) \approx 2 (\text{Ncm}^{-2}) \text{ at } f = 0.15. \quad (5)$$

Key Applications

For many years, traditional brushes for slip rings and commutators have either been a lubricated monofilament brush or a self-lubricating composite brush (Shobert 1965). These brushes have been successfully used, but improved performance is always desirable. Specially, today's industrial/commercial slip rings are required for high operational speeds, long life, no maintenance, and data transfer capability.



Metal Fiber Brushes, Fig. 2 An example of metal fiber brush (<http://www.moog.com/about/components>)

Metal Fiber Brushes, Table 1 Comparison of electrical brushes for space applications (<http://www.moog.com/about/components>)

Composite brushes	Monofilament brushes	Fiber brushes
Ag/MoS ₂ /graphite	Lubricated	Unlubricated
Most flight history	Considerable flight history	Growing flight history
Meets outgassing requirements	Does not meet outgassing requirements	Meets outgassing requirements
Self lubricating contacts (solid lubricant in brushes)	Requires liquid lubricant on contact surface	No oil or dry film lubricant application required
Requires largest ring-to-ring axial pitch	Requires smallest axial pitch (50% of composite brush pitch)	Requires less axial pitch than composites (70% of composite brush pitch)
Manufacture subject to greatest number of process variables	Manufacture subject to few process variables	Manufacture subject to few process variables
Greatest amount of wear debris generation (approximately 100 times the wear rate of fiber or monofilament brushes)	Much smaller amount of wear debris generation than composite brushes	Smaller amount of wear debris generation than composite and monofilament brushes
High electrical noise if operated in humid environment	Low electrical noise in air and vacuum	Low electrical noise in air and vacuum
Must operate in vacuum or dry inert atmosphere	Operational in air or vacuum with lubricant present	Operational in air or vacuum
Wide operating temperature range	Viscosity limited operating temperature range	Wide operating temperature range
Wide range of surface speeds	Limited range of surface speed	Wide range of surface speeds

Under the above circumstances, fiber brushes have been developed mainly for high current use and space application (Adkins and Kuhlmann-Wilsdorf 1979, 1980; Reichner 1980a, b). A chart comparing

the advantages of three types of brushes is shown on Table 1.

Metal fiber brushes are expected to have many advantages over conventional brushes in homopolar machine,

military, and aerospace applications as follows (Boyer and Chabrierie 1983):

- Multiple points of ring contact per brush bundle
- Ability to perform in ambient conditions as well as in vacuum conditions
- Contact surfaces that do not require lubrication
- Long life
- Low contact force per fiber
- Low contact wear rates (Boyer and Chabrierie 1982)
- High current density capability
- Low dynamic contact resistance (noise)

As an alternative to traditional sliding contact designs, fiber brushes have been used in many demanding applications such as:

- CT scan systems
- High-speed testing
- Robotic welding systems
- High-speed, in-line inspection systems
- Radar platforms
- Homopolar machines

In the future, fiber brushes may be used not only in high-tech applications but also in power stations, auxiliary car and airplane motors, and other industrial/consumer products, such as large industrial and marine motors, turbo generators, starter motors, hair dryers, and so on (Kuhlmann-Wilsdorf and Alley 1989; Brown et al. 2008).

Cross-References

- [Brush Materials](#)
- [Friction Coefficient](#)
- [Sliding Electrical Contact Wear](#)
- [Sliding Wear](#)

References

- C. M. Adkins III, D. Kuhlmann-Wilsdorf, Development of high-performance metal fiber brushes II- testing and properties, in *Electrical Contacts- 1979: Proceedings of the 25th Holm Conference on Electrical Contacts*, Illinois Institute of Technology, Chicago, IL, 1979, pp. 171–184
- C. M. Adkins III, D. Kuhlmann-Wilsdorf, Development of high-performance metal fiber brushes III- further tests and theoretical evaluation, in *Electrical Contacts- 1980: Proceedings of the 26th Holm Conference on Electrical Contacts*, Illinois Institute of Technology, Chicago, IL, 1980, pp. 67–72
- J.E. Archard, Contact and rubbing of flat surfaces. *J. App. Phys.* **24**, 981 (1953)
- L. Boyer, J.P. Chabrierie, J. Saint-Michel, Low wear metallic fiber brushes. *Wear* **78**, 59–68 (1982)
- L. Boyer, J.-P. Chabrierie, Current collection in a homopolar machine using metallic fiber brushes, in *Proceedings of the 1983 International Current Collector Conference*, Annapolis, MD, 1983, pp. 117–138

- L. Brown, D. Kuhlmann-Wilsdorf, W. Jesser, Testing and evaluation of metal fiber brush operation on slip rings and commutators. *IEEE Trans. Compon. Packag. Manuf. Technol.* **31**(2), 485–494 (2008)
- R. Holm, *Electric Contacts* (Springer, Berlin/New York, 1967)
- D. Kuhlmann-Wilsdorf, Electrical fiber brushes-theory and observations. *IEEE Trans. Compon. Packag. Manuf. Technol. A* **19**(3), 360–375 (1996)
- D. Kuhlmann-Wilsdorf, D. Alley, Commutation with metal fiber brushes. *IEEE Trans. CHMT* **12**, 246–253 (1989)
- D. Kuhlmann-Wilsdorf, Theoretical performance limits for different types of electrical brushes, in *Electrical Contacts- 1983: Proceedings of the 29th Holm Conference on Electrical Contacts*, Illinois Institute of Technology, Chicago, IL, 1983, pp. 21–30
- P. Reichner, Metallic brushes for extreme high current applications, in *Electrical Contacts- 1979: Proceedings of the 25th Holm Conference on Electrical Contacts*, Illinois Institute of Technology, Chicago, IL, 1980a, pp. 191–197
- P. Reichner, High current tests of metal fiber brushes, in *Electrical Contacts- 1980: Proceedings of the 26th Holm Conference on Electrical Contacts*, Illinois Institute of Technology, Chicago, IL, 1980b, pp. 73–76
- E.I. Shobert II, *Carbon Brushes* (Chemical Publishing, New York, 1965)
- P.G. Slade, *Electrical Contacts* (Marcel Dekker, New York, 1999)

Metal Forming

- [Friction in Metalforming](#)

Metal Removal Fluids (MRFs)

- [Cutting Fluids and Their Environmental Impact](#)

Metal Shaping

- [Friction in Metalforming](#)

Metal Working Fluids

- [Function of Cutting Fluids and Lubricants](#)

Metall-, Cermet- or Ceramic-Polymer-Composite Coatings for Tribological Applications

- [Self-Lubricating Treatment of Light Alloys](#)

Metallized Bound Carbon

► [Materials for Mechanical Seals](#)

Metalworking Lubricants

JAMES R. ANGLIN

Alcoa Technical Center, Alcoa Inc., New Kensington, PA, USA

Synonyms

[Cutting lubricants](#); [Forming lubricants](#); [Grinding lubricants](#)

Definition

Metalworking lubricants are products that are introduced between the tool and the workpiece to facilitate forming and metal removal processes. Their primary function is to control friction and limit tool wear while providing the desired surface quality to the workpiece. A wide range of compositions are commonly used, including fluids ranging from organic liquids and emulsions to glasses, and solids ranging from waxes to inorganic materials with layered structures well suited to lubricating.

Scientific Fundamentals

The functional requirements and formulation strategies for metalworking lubricants used for forming and for metal removal are discussed in this entry. The reader is referred to the entry entitled “[Chemistry of Rolling Lubricants](#)” as well as other entries listed under Cross-References, for further information in related areas. Laemmle, Mang et al., Passman and Schey may also be consulted for additional background.

Forming

In the forming process, lubricants are applied to control friction and wear at the interface between the metal and the tooling in hot and cold operations that include rolling, forging, extrusion, drawing, and a variety of sheet-forming applications. With the reshaping of the workpiece in a forming process, an amount of freshly exposed surface will be generated that will vary with the process conditions. A good lubricant can facilitate the production of complex shapes as faithfully as possible, with the boundary components serving to minimize adhesion

and surface disruption, especially where the new surface is exposed. In rolling, the roughness of the surfaces of the rolls, utilized to provide traction for the metal in the mill for most products, will also lead to a certain amount of abrasive wear with debris formation.

Metal Removal Processes

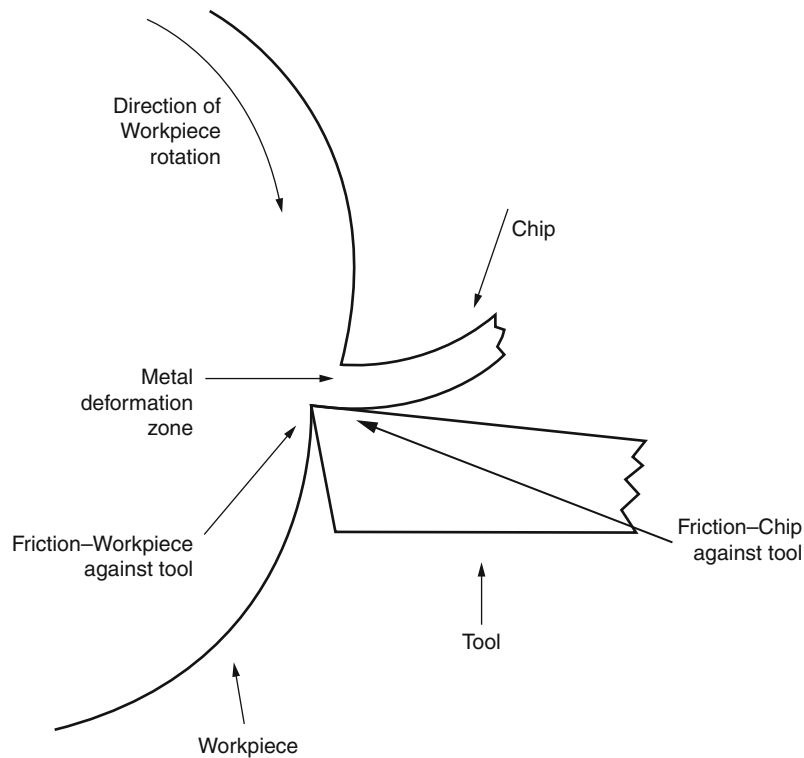
The removal of metal can transform a formed product into a more finished item. One can think of this as involving cutting processes and abrasive machining processes such as grinding. In general, these are performed at low rather than elevated temperatures.

Cutting

This category covers a range of processes, including drilling, tapping, turning, milling, and broaching. The tool can be thought of as being relatively wide and, as in [Fig. 1](#), a chip can be continuously generated as the cutting proceeds. The functions of the lubricant are to provide (1) cooling to the surfaces where heat is generated, especially under high-speed conditions; (2) lubrication to the contact areas, including the sliding of the chip over the front (rake) surface of the tool as well as the sliding of the bottom (flank) surface of the tool over the freshly cut surface; and (3) flushing of chip fragments from the cutting zone. The lubricant can contribute to improved tool life by controlling the adhesion of freshly created surface to the tool and minimizing thermal shock associated with the cutting process. Adequate coolant flow well directed to the contact areas will provide the cooling and also serve to flush debris away from the work zone.

Abrasive Machining

In this process, metal is removed using abrasive media. The process can range from creep-feed grinding, where the goal is rapid metal removal, to more gradual metal removal processes including honing and lapping. The grinding media are hard materials of suitable size and friability, such as alumina and silicon carbide, and are bonded into a matrix in a shape such as a wheel. Superabrasives such as cubic boron nitride and diamond can similarly be bonded into grinding tools. The abrasive may alternatively be attached to a flexible backing such as paper or a belt. In addition, the abrasive may be incorporated into a slurry and metal removed with the aid of a suitable tool, as in lapping; it may also be blended into a semisoft mixture and buffed against the workpiece using a soft material such as a fabric. Abrasive machining can be combined with electrolytic effects in electrochemical grinding and chemical effects in chemical mechanical polishing.



Metalworking Lubricants, Fig. 1 Illustration of cutting mechanism

M

In the grinding process, performance will depend on the ability of the wheel or other device to continue to provide consistent performance over time. This can include the ability of the abrasive to be friable and regenerate good cutting surface, and also for the composition to gradually release spent or worn abrasive, exposing fresh abrasive material.

A number of processes take place in the grinding zone, where the nature of the numerous contacts between abrasive and workpiece vary according to the size, shape, and orientation of the individual abrasive particles. Among these is a cutting mechanism, where the grit penetrates to a sufficient depth and has a shape and orientation such that a chip will form. For some abrasive particles, the contact does not favor chip formation, whether because of orientation, wear pattern, or buildup of wear debris, and the resulting ploughing contact generates a furrow, as shown in Fig. 2, displacing the metal but not removing it. Rather than having chip formation, a prow will build up at the leading edge of the tool. In addition, some abrasive particles with very shallow penetration will lead only to elastic deformation, with frictional heating being the main result.

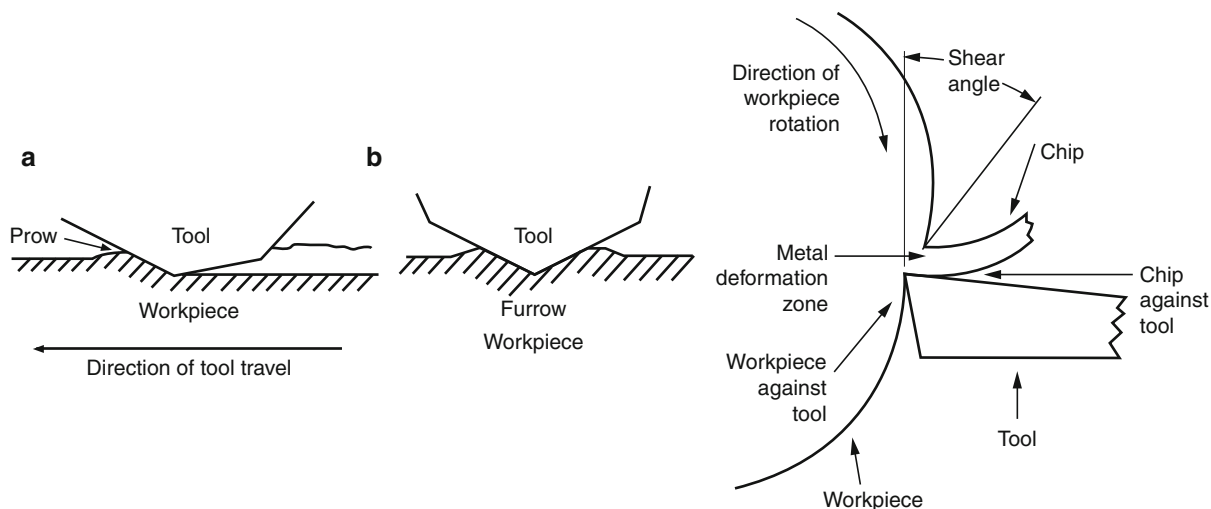
In abrasive machining, sliding friction is a key component of heat generation and much of this heat remains in the workpiece, rather than being conveyed away with the chip, as in the cutting process. Thus, the reduction of friction is important in abrasive machining. The fluid choices and formulation options remain similar to those for cutting and other metalworking operations.

Key Applications

Lubricant Requirements

As with most lubricants, the requirements that a metalworking lubricant will need to meet extend beyond suitable control of friction and wear. The relative importance of each requirement and the performance of the lubricant against them will lead to the choice of the best product for a given situation. Some of the important requirements of lubricants, all of which will be discussed in this entry, include:

- Good product surface finish (absence of scuffing, galling, fissures, stain, and hard-to-clean residues)
- Good tooling life
- Good cooling performance



Metalworking Lubricants, Fig. 2 Illustration of ploughing mechanism panel (a): side view showing prow at leading edge; panel (b): end view

- Suitable environmental, health, and safety performance
 - Ease of waste treatment
 - Acceptable level of volatile organic compounds
 - Minimal mist generation
 - Absence of toxic and carcinogenic components
 - Minimal tendency for skin irritation
 - Acceptable odor
- Robust formula providing consistent performance for an extended time over a range of processing conditions
 - Tolerates or rejects lubricants that can leak into it or commingle with it
 - Debris is readily removed and does not destabilize the formulation
 - Resistant to microbial attack
 - Key components can be monitored and added as needed
 - Easily applied uniformly and consistently
- Commercially suitable
 - Compliance with applicable regulatory requirements, e.g., U.S. FDA
 - Use of components from renewable resources
 - Low cost
 - Equivalent products from multiple reputable suppliers
- Absence of corrosion of tooling, workpiece, and equipment surfaces

Common Lubricant Types

Products based on hydrocarbon oils or other nonaqueous liquids have a long history of use and include a wide range

of formulations. The hydrocarbons can range from linear paraffins and hydrotreated kerosenes used in cold rolling of aluminum to much heavier products, including synthetics, that can be used in various forming operations including cold rolling of steel, hot rolling, forging, and extrusion. Among the synthetic hydrocarbons, polyalphaolefins can provide enhanced high and low temperature performance, and polyisobutylenes that are available in a wide range of viscosities have good resistance to stain because of their tendency to depolymerize at higher temperatures. Esters from vegetable and animal sources can replace hydrocarbon oils where the use of renewable resources is important. These fluids as well as synthetic esters typically have good inherent lubricity and high flash points. However, anhydrous fluids do not have the capacity for cooling that aqueous fluids provide. Additives can offer enhanced anti-wear and boundary performance, rust and oxidation protection, foam and mist control, and other properties as needed. The boundary additive chemistry is tailored to the metal being worked. Phosphorus, chlorine, and sulfur compounds are commonly used for steel, with sulfur having the best performance under high temperature conditions. Sulfur can be problematic, however, for some metals, and care needs to be taken to avoid organochlorine compounds with known health risks. Oleochemical products are typically used for boundary lubrication with aluminum and copper alloys. These can include natural oils and fats, fatty acids, fatty alcohols, and synthetic esters. Oil-based fluids can be simpler to maintain than water-based fluids since concerns associated with water, such as the control of emulsion concentration and

stability, water quality, corrosion, and microbes, are avoided.

Emulsions of the oil-in-water type provide one means to introduce the enhanced cooling capacity and fire resistance associated with the presence of water. The misnomer *soluble oil* has been used for these fluids, which are similar to the ISO HFA-E class for hydraulic fluids. A common formulation, as may be used to lubricate machinery, can consist of naphthenic oil emulsified with an anionic emulsifier, such as a natural or synthetic sulfonate, and contains rust inhibitor and E.P. additives. These fluids typically have an appearance similar to milk (they scatter white light) and have droplet sizes greater than about 0.5 μm . More broadly, this class can extend to emulsions containing multiple emulsifiers and a variety of boundary additives, and can also include emulsified vegetable and animal oils. They are widely used in metalworking and in rolling, with the additive chemistry tailored to the metal of concern.

The stability of the emulsion is important, since the fluid's tribological performance requires sufficient oil delivery to the tool-workpiece contact. Higher concentrations and larger droplet sizes favor optimum lubricant delivery, and so a less stable emulsion with larger droplet sizes is capable of providing better performance. However, this has to be balanced against the greater difficulty in ensuring that the oil remains evenly distributed in the water with a consistent droplet size. One potential drawback with emulsions is sensitivity to contaminant oils, since the introduction of a given volume of contaminant can have greater impact when the total amount of oil present in the form of an emulsion is relatively small. Contaminating oils, if not of the same composition as the metalworking fluid, can be skimmed off the surface in a reservoir or sump, but avoiding some entrainment of the oil in the fluid or some partitioning of the metalworking composition into the contaminant before skimming is difficult, and so performance is usually degraded.

The quality of the water used for blending and for replenishment of evaporated water is important. If the water is not deionized, cations such as calcium in hard water can react with anionic surfactants and with fatty acids in the formulation, converting them to poorly soluble soaps and reducing their effectiveness. At the same time, chloride in hard water can build up to levels that present a significant corrosion concern. The overall increased conductivity can also reduce emulsion stability. The presence of phosphate and sulfate in the water can encourage microbial activity.

The selection and control of emulsifiers is a key to the performance of oil-in-water emulsions. For robust

performance, it is often advantageous to use multiple emulsifiers, including, for example, nonionic and anionic types, and to do so using a suitable combination of their HLB (hydrophilic-lipophilic balance) values. These reflect the relative affinities of the emulsifiers for the water and oil phases and can be selected to provide optimum surfactant efficacy. Many emulsifiers are complex mixtures that are difficult to monitor by commonly available methods, and so controlling their content is difficult. Some are also built into molecules that have other functions, such as boundary lubricant performance. As a result, replacement of part or all of the fluid volume with a fresh blend may be necessary when performance degrades.

With high levels of a suitable surfactant blend and reduced levels of oil, thermodynamically stable, translucent microemulsions can be formed that have quite low interfacial tension. These fluids are frequently referred to as semi-synthetics and can have a bluish color with droplet sizes typically in the 0.01–0.2 μm range. Excess surfactant may be present in the form of micelles in the fluid. With the decreased availability of the oil because of the small droplet size, these products may be favored in applications where the emphasis is on cooling more so than tribological performance.

When there is no oil content, transparent solutions of surfactants and such additives as corrosion inhibitors can be formulated. These are frequently termed synthetic products. Here there are no oil droplets in the fresh solutions, although surfactant micelles can be present. Somewhat similar are solutions containing polyglycols that are water soluble and used in some applications, such as rough machining. Like microemulsions, the solution products are very good coolants and they can be expected to have good resistance to microbial attack. With the high surfactant content, foaming may be a concern. While defoamers can be helpful, care needs to be taken in their selection since traces of silicone defoamers can be very detrimental to the adhesion of any coatings applied in downstream processes. These synthetics can tend to emulsify oils that leak into the system or to leach lubricants from equipment they contact, leading potentially to more rapid wear in that equipment. For solutions with high levels of glycols, attention may be needed in the choice of paints and elastomers in the system to ensure they are compatible with these fluids.

An emerging trend is the use of minimum quantity lubrication (MQL), which is the use of very small quantities of lubricant delivered as an aerosol to the metalworking application, particularly metal removal applications. Where successful, MQL can minimize the issues with handling and disposal of potentially quite large volumes

of lubricant. It may also enhance tool life, perhaps by minimizing the temperature variations and thermal shock compared with what the tooling experiences in intermittent cutting operations under constant coolant flow. It will have limitations in terms of heat removal, chip and swarf removal, and potential inhalation of the mist needs to be considered. However, it is a significant step towards the concept of machining without lubricants (dry machining). With careful selection of the process and metal as well as the tooling, dry machining has shown some promise. With dry machining, the potential hazards of finely divided reactive metal debris need to be recognized.

Lubricants for Metal Forming Processes

Many forming processes are carried out successfully at room temperature and use only organic media as the lubricant. These media can range from hydrocarbon oils with additives tailored to the workpiece composition to formulations based primarily on oleochemical materials. Waxes or polymer coatings or films may also be used and can provide low friction, although hydrocarbon waxes do not provide good wear protection. In some processes, such as the forming of aluminum lids for cans, a very thin layer of lubricant on the surface of a polymeric coating lubricates sufficiently to facilitate forming. However, some operations, notably those at high speed where cooling is important, use water-based lubricants. Examples include the cold rolling of steel and the drawing and ironing process for making beverage cans.

If interface temperatures in the forming process are 400–450°C or higher, organic materials become less able to avoid decomposition and the resulting loss of their effectiveness, especially if the contact times are more than very brief. The burning of oil-based lubricant remains common in high-temperature forging of aluminum. Extensive lubricant decomposition can lead to adhesion between the die and workpiece with the result being scuffing or galling of the workpiece. The use of water-based lubricants such as emulsions can provide real benefits in this situation, and so hot rolling of aluminum and steel is always performed with water-based products. Where the time and temperature conditions of the contact are too severe for an organic fluid alone, the addition of lamellar solid lubricants, such as graphite, hexagonal boron nitride, molybdenum disulfide, and tungsten disulfide, to the fluid may provide important benefits. Similarly, benefits can come from the use of soluble metal derivatives, such as carboxylate salts of metals such as zinc, tin, and lead that spread readily with the oil base and can be expected to decompose to the solid oxides. Lead, however, has largely been abandoned for health

reasons. The use of solid lubricants as a loose powder is challenging as it is difficult to obtain good adhesion to the die or workpiece. Without adequate adhesion, the solid lubricant used, for example, in superplastic forming processes may simply be pushed aside, leading to scuffing in the resulting bare areas and undesired accumulations in other areas. For higher temperature operation, the use of inorganic materials that melt, such as glasses, is an option and is utilized in the Sejournet process for high-temperature extrusion. Simple use of a layer of thermally resistant material in high-temperature forming is an option, but such an approach sacrifices some fidelity to the desired shape in order to minimize potential sticking to the die.

Lubricants for Metal Removal Processes

In general, cutting and grinding operations are performed at low rather than elevated temperatures and the lubrication options cover the range of oil- and water-based lubricants as well as MQL and dry machining. The choice of the fluid type and composition will depend on the specific process conditions, including the relative importance of friction reduction and cooling, together with the metallurgy of the tooling and workpiece.

Lubricant Application

The application methods for metalworking lubricants are nearly as diverse as the products themselves. Electrostatic application methods can be used for uniform application at high speed of sheet forming lubricants at levels as low as 50 mg/m², or to apply solids or suspensions of solids to surfaces prior to high-temperature forming, with computer control where required. Equipment to generate trace amounts of atomized liquids in air is used for MQL. In forging applications, rapid application of the lubricant formulation in the correct amounts to the necessary areas of the dies is needed for high productivity. For many applications, flooding of the contact occurs, with an extreme example being the thousands of liters per minute of coolant applied from computer-controlled networks of nozzles in the rolling process. In all instances, this needs to be accomplished with minimal release of mist or overspray outside the process area.

Fluid Maintenance Considerations

For consistent performance, it is crucial that the composition of the lubricant being applied is consistent and correct. While cost is always a factor, it is important to use quality controlled product from reliable suppliers and, where possible, to select products available from multiple sources. It is important to avoid any contamination or degradation prior to use at the manufacturing site.

This can include the separation of emulsion products resulting from freezing or the degradation of products under poor or extended storage. For systems where a fluid is recirculated, important factors include the control of contamination by other fluids and maintaining the quality of the water phase in water-based products.

For large systems containing lubricant volumes whose replacement value can exceed US\$100,000, it is important to be able to monitor and adjust composition to minimize the need to replace or reclaim a significant part of the volume. The compositional monitoring is typically performed by chromatography and spectroscopy.

In operations such as rolling and metal removal, the control of the debris formed is crucial to producing metal of suitable surface quality. The performance of the product should not be adversely affected by typical debris levels. While a number of control options are available, filtration is most commonly used. The use of filter aids such as diatomaceous earth and fuller's earth is typically needed if fine debris need to be removed. This tends to be costly in terms of labor, disposal expenses, and the cost of lubricant disposed of together with the filter media. In addition, it needs to be realized that fine debris as formed in the absence of water can present an ignition hazard.

An additional consideration for water-based formulations is microbial control, particularly for emulsion products. A number of strategies are available, ranging from operating at temperatures of $\sim 60^{\circ}\text{C}$ and above to control growth, using components with some resistance to microbial attack, and employing ultrasound and UV light. Nevertheless, in practice, it proves difficult for such approaches to be fully successful.

The use of biocides is, therefore, a common practice for water-based systems, and biocide additions are made based on the results of tests of microbe population. The choice of a suitable material, or preferably a pair of materials so that the second can be used to control microbes developing a tolerance for the first one, does require some care. The various candidate materials vary in their rate and duration of action and in their relative performance against bacteria, yeast, and mold. Some can react with chemical species found in metalworking formulations, affecting their performance, and some can react with other types of biocides.

Lubricant Handling Considerations

A number of potential hazards exist in the handling of metalworking lubricants, including exposures to chemicals through skin contact and inhalation of mist and organic vapors, which can also present an environmental hazard. Both the formulator and user share in the

role of ensuring that the products in use have acceptable hygiene properties. This includes the avoidance of components with potential toxic or carcinogenic properties and tendencies for skin irritation. The concern extends to the odor of the formulation, which is important to plant personnel. Nevertheless, proper use of personal protective equipment and engineering controls to minimize worker exposure is important. Biocides must be handled with care when concentrated; however, it is noted that some are routinely used at low levels in personal care products.

Reclamation and Disposal

The minimizing of waste lubricant is an important goal of sustainable production. As mentioned above, control of composition and avoidance of contamination are two keys to enable continued satisfactory usage of a product. However, byproducts and contaminants will inevitably build up, putting the responsibility on the formulator to develop products that can be reclaimed and can be handled successfully with available waste treating options.

With oil-based formulations, the low-viscosity products used in cold rolling of aluminum can be formulated such that the base oil and additives can be efficiently separated from hydraulic oil and other higher molecular weight materials by distillation, preferably under vacuum. Such an approach, however, is more difficult with the higher viscosity products used in steel rolling and aluminum hot rolling. Separated oils, if no longer suitable for metalworking applications, may then be burned to recover their energy value. Used filter media containing significant amounts of entrained lubricant oils can potentially be treated to recover the oil for reuse. Alternatively, the media can be shipped to a suitable facility to recover the energy value of the oil and utilize the residual solid, avoiding the alternative of disposing of it in a landfill.

For water-based formulations, the ease of reclamation will depend on the types and amounts of emulsifier, with nonionic ones being the most challenging to treat chemically. In some situations, the simplest approach may be removal of the water by distillation with recovery of the oil for its energy value. The recovered water may then be reusable in the same process or other plant applications. Another approach is concentration of the oil phase by ultrafiltration, thereby reducing the volume of oily waste and generating water of very low oil content. In addition, the use of strong acid can aid in the separation of oil from water, as can the use of various flocculating agents. The selection of the best technique or combination of techniques needs to be made in keeping with current and foreseen water quality regulations.

Acknowledgment

The author acknowledges the helpful assistance of Prof. S. R. Schmid in the preparation of this manuscript.

Cross-References

- [Chemistry of Rolling Lubricants](#)
- [Function of Cutting Fluids and Lubricants](#)
- [Lubrication with Emulsions](#)

References

- J.T. Laemmle, in *Metalworking Lubricants*, ASM Handbook, vol 18 (1992), pp. 139–149
- T. Mang, C. Freiler, D. Hörner, Metalworking fluids, in *Lubricants and Lubrication*, 2nd edn. ed. by T. Mang, W. Dresel (Wiley-VCH, Weinheim, 2007), pp. 384–521
- F.J. Passman, Microbiology of metalworking fluids, in *Metalworking Fluids*, 2nd edn. ed. by J.P. Byers (CRC Press, Taylor & Frances Group, Boca Raton, 2006), pp. 195–229
- J.A. Schey, *Tribology in Metalworking* (American Society for Metals, Metals Park, 1983), pp. 573–688

Methods of Matched Asymptotic Expansions in Heavily Loaded EHL Contacts

- [Asymptotic Methods for Analyzing Heavily Loaded EHL Contacts](#)

Micro Deep-Drawing

- [Friction in Microforming](#)

Micro Extrusion

- [Friction in Microforming](#)

Micro Stamping

- [Friction in Microforming](#)

Micro Vibrations

- [Contact with Micro Vibrations for Friction Control and Wear Reduction](#)

Micro-arc Oxidation (MAO)

- [Ceramic Conversion of Light Alloys](#)

Microcracks

- [Growth Characteristics of Small Fatigue Cracks](#)

Micromechanics for Contact Applications

LEON M. KEER, XIAOQING JIN

Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA

Synonyms

[Eigenstrain method](#)

Definition

The continuum mechanics theory related to microstructures of materials at the scales of microns or less. In particular, problems relating to inclusions and dislocations are most effectively analyzed by this method.

Scientific Fundamentals

The contact of bodies is usually taken at the scale of the theory developed by Hertz (1882a, b). For that theory frictionless, smooth contact was assumed. The contacts, although small, were usually of the scale of millimeters. However, as steels became cleaner, the surface topography became an important contributor to component failure, compared with the deep failures associated with Hertz. Therefore, the scale shifted to that of materials science, which is on the order of microns or less. The mechanics of inhomogeneities and grains became important in

predicting the life of bearings and gears. The term micromechanics refers to studies at the scale of material microstructure. As pointed out by Mura (1991), “the method employed is a continuum theory of elasticity; yet its applications cover a broad area relating to the mechanical behavior of materials: plasticity, fracture and fatigue, constitutive equations, composite materials, polycrystals, etc.” Methods used by Mura are based on the “eigenstrain method,” where eigenstrain is a generic name that refers to stress-free transformation strains, such as thermal expansion, phase transformations, and misfit strains. J.D. Eshelby (1957, 1959) was a pioneer in this field and coined this term in revolutionary papers. For further information concerning Eshelby’s contributions, one is referred to the collection of his papers edited by Markenscoff and Gupta (2006).

This section will consider two examples of how theories of micromechanics can give predictive results for determining the life of such components when subjected to contact loading.

Key Applications

Problems Related to Three-Dimensional Inclusions in a Half Space

The method developed by Eshelby (1957, 1959) was applied to an ellipsoidal inclusion in an infinite space. Uniform eigenstrains were assumed in this region. The solution depended upon the integration in an ellipsoidal region of the point load solution for an infinite space, Kelvin’s solution. For uniform eigenstrains, Eshelby found that the interior strain field and therefore the stress field were also uniform. This amazing result was expressed as a linear relation between the interior strain field and the interior eigenstrains. This relation is called Eshelby’s tensor and it is given in terms of complete elliptic integrals. In case the ellipsoid is a sphere, then the tensor becomes a result in a closed form. Of the two Eshelby papers, the first gave the interior field and the second the exterior field. The solution was especially useful in the field of materials science as the ellipsoidal shape is fairly general (it can represent cracks, precipitates, inhomogeneities) and there are many applications.

Another geometry that has been resolved in a closed form is that of an inclusion in an infinite elastic medium with uniform eigenstrain. This problem, solved by Y.P. Chiu (1977), provides a building block for interaction type solutions involving inclusions that do not follow the assumptions of Eshelby or Mindlin and Cheng.

However, for contact applications, the closed form results of Eshelby are of limited use and require intensive

numerical computations. The problem arises because contact applications require solution of an inclusion beneath a free or loaded surface. If more than one inclusion is considered, then the computational penalty becomes severe. The problem was brought to a more reasonable computational time scale through efforts Jacq et al. (2002), who developed a three-dimensional semi-analytical elastic–plastic contact code. The group of tribologists at INSA-Lyon have also produced several additional papers, advancing the solutions to include, thermal effects, elastic–plastic contact, as well as others. Due to their reliance on Chiu’s solutions as well as the fast Fourier transform (FFT) algorithm, the method is termed the semi-analytical method (SAM).

It appears that the approach can be made even more computationally friendly through use of a highly efficient algorithm to achieve two goals: a finer discretization and a larger number of computational steps. A recent paper by Zhou et al. (2009) solves the inclusion problem through utilization of a combination of 3D FFT and 2DFFT algorithms. Again the solution depends on the use of the SAM and solution by Chiu (1977). Solution of the inclusion problem implies that similar techniques together with Eshelby’s equivalent inclusion method can be used to realize the inhomogeneity problem.

Theory of Contact Fatigue as Dislocation Buildup

The formation of an “engineering” size (e.g., 3–10 mm) fatigue crack has been conventionally represented by two distinguishable stages: initiation and propagation. In the initiation stage, the damage accumulation occurring on the scale of the crystal lattice, dislocations, grains, etc., originates the nuclei of the microstructurally small cracks (MSCs) (Suresh 1998, p. 543). The subsequent propagation stage is characterized by the growth of the mechanically small cracks. The growth of MSCs is very sensitive to local grain properties. As its size becomes greater than the critical length of MSC, the fatigue crack transits to the mechanically small crack, whose propagation exhibits little or no pertinence to microstructure.

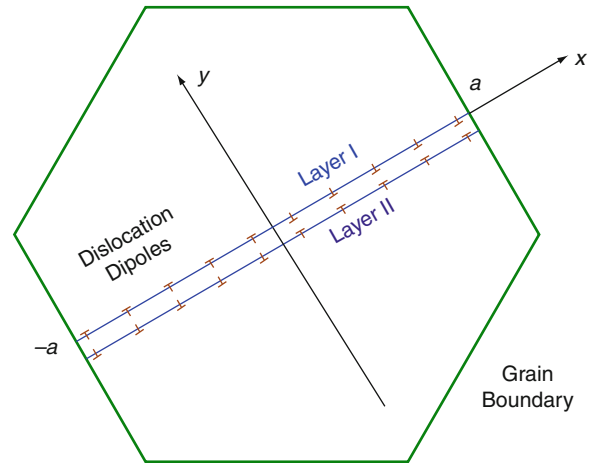
Historically, fatigue theory presented at the early stage was described using an *S-N* diagram, relating stress or strain (*S*) and cycles-to-failure (*N*). Lundberg and Palmgren (1947) (L–P) gave the probability of survival of an elemental volume loaded by a nonchanging cyclic stress. Since fatigue failure generally requires the establishment of a cyclic plastic zone in the bulk material, Coffin (1954) and Manson (1954) (C–M) proposed that the fatigue life of a material specimen is determined by the cyclic plastic strain amplitude through a power law relation. In practice,

a number of major achievements deploying the S - N fatigue tests have been successfully applied to engineering design. However, due to a lack of the detailed microstructural information in these semi-phenomenological models, they are incapable of predicting the MSCs stage during the early evolution of fatigue cracks. Moreover, since the final conditions of the fatigue cracks nucleation and early growth become the initial conditions for the standard fatigue tests, systematic investigation of the fatigue crack initiation from a micromechanical point of view integrated to the classical fatigue theory, facilitates the adaptability of both the experimental and theoretical explorations.

The initiation of fatigue cracks is a complicated phenomenon involving various micromechanical interactions among dislocations, grains, intrinsic resistance force, etc. Before they nucleate into macrocracks, the fatigue cracks primarily develop in the patterns of shear cracks that find the easiest way to propagate. The site of crack initiation varies depending on both the microstructure and the stress state of the material under investigation. A large number of metallographic observations have indicated that the localized dislocations pile-up in persistent slip bands (PSBs) is an important mechanism responsible for fatigue crack initiation in many materials.

It is well known that dislocations can be viewed from the point of view of eigenstrain theory mentioned earlier. Tanaka and Mura (1981) developed a theory to explain the ratcheting of the localized plastic deformation along the PSBs under a cyclic loading. Their micromechanical model is based on the idea that there are two adjacent layers of dislocation pileups on a grain located in the specimen, with one part sliding during forward loading and the other sliding during reverse loading (Fig. 1). The fundamentals of the dislocation dipoles model agree with the observations in the experiments (cf. Suresh 1998, p. 140) that there exists two closely spaced slip planes, with one accommodating plastic deformation during the forward loading and the other being associated with the reversed loading.

By utilizing the theory of continuously distributed dislocations, Tanaka and Mura are able to calculate the dislocation density function, plastic strain, and the associated energy during each loading cycle. Consequently, a systematic buildup of accumulated dislocation dipoles is achieved under the assumption of irreversible dislocation motion in the favorably oriented grain. This model is one of damage accumulation and gives scientific validity to the phenomenological Coffin–Manson equation. Tanaka and Mura (1982) later extended this theory to fatigue crack initiation at inclusions. Assuming a circular



Micromechanics for Contact Applications, Fig. 1 Mura's model for crack initiation within a most favorably oriented grain

cylindrical inclusion embedded in the matrix subjected to a uniform anti-plane shear stress, they analyzed three different mechanisms for fatigue crack initiation: (1) The inclusion is debonded from the matrix by the first loading, and the inclusion becomes mechanically equivalent to a stress-free pore; (2) The dislocation accumulated in the matrix impinges on the inclusion, eventually resulting in the inclusion debonding or cracking; (3) The fatigue crack initiates at the slip band emanating from an uncracked inclusion.

The crack initiation theory was later extended by Mura and Nakasone (1990), who developed an analogy to Griffith's theory of fracture that applied to fatigue. The Gibbs free energy change from a state of dislocation dipole accumulation to that of a crack of size c is given by

$$\Delta G = -W_1 - W_2 + 2c\gamma_s \quad (1)$$

where γ_s is the surface energy of each crack face, W_1 the elastic strain energy built up in the material due to the dislocation dipoles accumulated in the PSBs, and W_2 the mechanical energy release. The stored energies W_1 and W_2 will be released when a crack forms but the surface energy term is necessary to create the two new crack faces. The calculation of Mura and Nakasone show that the Gibbs free energy change in (1) is a quadratic function of the number of loading cycles n . The system becomes unstable when ΔG takes a maximum value at $n = n_i$. It is postulated that n_i is the critical cycle number for fatigue crack initiation under a given cyclic load.

Bhat and Fine (2001) examined (1) from the standpoint of phase transformation theory, which is essentially

a thermodynamic approach. Fatigue crack initiation is considered as “a nucleation process due to random fluctuation in a metastable assembly of defect structures generated during cyclic deformation.” During the formation of a fatigue crack, there is an energy barrier due to the energy required to create new surfaces. On the one hand, the accumulated damage in the form of increased lattice defects due to cyclic loading increases its internal energy. On the other hand, when a crack forms in a solid material, defects are released as well as the elastic energy from the applied load.

Based on Fine’s theory (Fine 2000), the increase in internal energy from damage accumulation is a measurable quantity that can be investigated experimentally. An energy density function, δ , defined as the stored energy in defects that are lost per unit area of the nucleated crack surface is introduced. Therefore, the first term on the right hand side of (1) becomes

$$W_1 = A\delta \quad (2)$$

where A is the area of just nucleated crack. The energy per cycle due to the accumulated defects is related to the hysteresis loop parameters as:

$$\frac{d\delta}{dN} = f\Delta\sigma\Delta\varepsilon_p \quad (3)$$

where $\Delta\sigma$ and $\Delta\varepsilon_p$ are the stress and plastic strain ranges in the cycle, and the efficiency factor f is taken as a constant representing the energy stored for each cycle. The importance of f also indicates the fact that most of the irreversible energy is converted to heat and only a small portion is stored as defects.

As a concluding remark, crack nucleation has received considerable attention in fatigue research over the years. However, the fundamental physical basis for this research still requires further study. The dislocation dipole model proposed by Mura and his associates tends to be a reasonable theory which provides a systematic point of view based on micromechanics. Fine proposes a science based equation by including measurable quantities related to microstructural information.

References

- S.P. Bhat, M.E. Fine, Fatigue crack nucleation in iron and a high strength low alloy steel. *Mater. Sci. Eng. A Struct. Mater. Prop. Microstruct. Process.* **314**(1–2), 90–96 (2001)
- Y.P. Chiu, Stress-field due to initial strains in a cuboid surrounded by an infinite elastic space. *J. Appl. Mech. Trans. ASME.* **44**(4), 587–590 (1977)
- L.F. Coffin, A study of the effects of cyclic thermal stresses on a ductile metal. *Trans. ASME.* **76**, 931–950 (1954)
- J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **241**(1226), 376–396 (1957)
- J.D. Eshelby, The elastic field outside an ellipsoidal inclusion. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **252**(1271), 561–569 (1959)
- M.E. Fine, Phase transformation theory applied to elevated temperature fatigue. *Scripta Materialia.* **42**(10), 1007–1012 (2000)
- H. Hertz, Über die Berührung fester elastischer Körper. *J. reine angewandte Mathematik.* **92**, 156–171. (English transl. H. Hertz, (1896) *On the contact of elastic solids. In: Miscellaneous Papers by H. Hertz.* D.E. Jones and G.A. Schott (eds.), Macmillan, London, 146–162, 1882a/1896)
- H. Hertz, Über die Berührung fester elastischer Körper und über die Harte. *Verhandlungen des Vereins zur Beförderung des Gewerbefleißes*, Berlin (1882). (English transl. H. Hertz, (1896) *On the contact of elastic solids and on hardness. In: Miscellaneous Papers by H. Hertz.* Eds. D.E. Jones and G.A. Schott (eds.), Macmillan, London, 163–183, 1882b/1896)
- C. Jacq, D. Nelias, G. Lormand, D. Girodin, Development of a three-dimensional semi-analytical elastic-plastic contact code. *J. Tribol. Trans. ASME.* **124**(4), 653–667 (2002)
- G. Lundberg, A. Palmgren, Dynamic capacity of rolling bearings. *ACTA Polytech. Mech. Eng. Ser.* **1**(3), 5–50 (1947)
- S.S. Manson, Behavior of materials under conditions of thermal stress. NACA Report 1170. Lewis Flight Propulsion Laboratory Cleveland, OH, 317–350 (1954)
- X. Markenscoff, A. Gupta, (Eds.), *Collected Works of J.D. Eshelby* (Springer, Dordrecht, 2006)
- T. Mura, *Micromechanics of Defects in Solids* (Kluwer, Dordrecht, 1991)
- T. Mura, Y. Nakasone, A theory of fatigue crack initiation in solids. *J. Appl. Mech. Trans. ASME.* **57**(1), 1–6 (1990)
- S. Suresh, *Fatigue of Materials* (Cambridge University Press, Cambridge, 1998)
- K. Tanaka, T. Mura, A dislocation model for fatigue crack initiation. *J. Appl. Mech. Trans. ASME.* **48**(1), 97–103 (1981)
- K. Tanaka, T. Mura, A theory of fatigue crack initiation at inclusions. *Metall. Mater. Trans. A* **13**(1), 117–123 (1982)
- K. Zhou, W.W. Chen, L.M. Keer, Q.J. Wang, A fast method for solving three-dimensional arbitrarily shaped inclusions in a half space. *Comp. Method Appl. Mech. Eng.* **198**(9–12), 885–892 (2009)

Micro-pitting

- [Rolling Bearing Fatigue Life, Effect of Profiles, Effect of Surface Roughness, Effect of Residual Stress](#)

Microstructurally Small Cracks

- [Growth Characteristics of Small Fatigue Cracks](#)

Microstructuring

- [Laser Texturing, Characterization and Related Effects](#)

Microtexturing

► [Laser Texturing, Characterization and Related Effects](#)

Microtribodynamics of Magnetic Storage Hard Disk Drives

RAJA R. KATTA, ANDREAS A. POLYCARPOU

Department of Mechanical Science and Engineering,
University of Illinois at Urbana-Champaign, Champaign
Urbana, IL, USA

Synonyms

[Adhesive contact of elastic bodies: the JKR theory](#); [Contact recording](#); [Friction/vibration interaction](#); [Head-disk interaction](#); [Nanomechanics](#); [Near-contact recording](#)

Definition

Microtribodynamics is the study of adhesive frictional contacts coupled with system dynamics in miniature systems.

Scientific Fundamentals

Contact

Contact is said to occur when two bodies near each other (whether sliding or not) “touch.” The study of contact is important because there are forces generated due to the contact of the bodies. These forces can in turn cause unwanted stresses in the bodies and result in damage.

Smooth vs. Rough Contact

The type of the surfaces of the bodies involved in contact is very important. The generated contact stresses depend on the geometry of the surfaces of the bodies in contact. Also, to some extent, all surfaces inherently possess a topography that is not mathematically flat but has some degree of irregularity. This is termed surface topography (or surface morphology), which in turn can be decomposed into waviness (larger spatial frequency components) and roughness (higher spatial frequency components). Note that topography may also include error-of-form (some form of slope). Typically the surface roughness dictates the magnitude of the resulting forces due to contact. In micro and nanoscale contact

applications, the effect of roughness on the contact can be more critical.

Friction

Friction can be either static friction (also referred to as stiction) or kinetic friction (also referred to as dynamic friction). Static friction is defined as the force just sufficient to prevent relative motion between two bodies, and implies that static friction is a constant value and can easily be divided by the external load to give the static friction coefficient. On the other hand, kinetic or dynamic friction force is the minimum force needed to maintain relative motion between two bodies. This definition implies that the dynamic friction and dynamic friction coefficient are time varying, even though they are typically reported as constant values.

Adhesion

Adhesion near contact and during contact arises from intermolecular forces acting between the two surfaces. With contacting and sliding surfaces becoming smoother and their size shrinking to micrometer and even nanometer scales, intermolecular interactions, known as adhesion between the interacting surfaces, become significant. In the literature, there are two main interpretations and understandings (approaches) of adhesion. The most common is adhesion, which is viewed as force acting between two surfaces, its physical origin being intermolecular, electrostatic, or capillary forces. The second approach is to express adhesion in terms of “thermodynamic work or free energy of adhesion,” which is the work or energy to separate unit areas of the interface. For micron size surface roughness and for heavy loaded contacts, adhesion force is usually negligible. However, for contacts in systems in the micro or nanoscale, the adhesive forces are important and need to be accounted for.

Key Applications

Magnetic Storage Hard Disk Drive Technology

In a magnetic storage hard disk drive operation, the information is stored as magnetic bits in a thin layer of magnetic material present in the rotating disk. A slider with the read/write recording head flying over the disk, as illustrated in [Fig. 1](#), senses the magnetic field from these bits. The data areal density increases with a reduction in the bit size. However, the magnetic field from these bits decays rapidly away from the magnetic material with the distance of the decay scaling with the bit size. Thus, as the

areal density increases, the spacing between the head and the disk must be further reduced.

The recording slider flies over the disk rotating at typical linear velocities of 5–50 m/s with a physical spacing (clearance) of a few nanometers. Certain features are precisely fabricated onto the bottom surface of the slider to enable it to fly using an air bearing. The air bearing is generated by the disk pulling air underneath the slider. The fabricated surface features form an air bearing surface (ABS) that generates a lifting force (F_a), which balances the loading spring force (F) from the suspension, enabling the slider to fly over the disk. The disk and slider surfaces are coated with a hard material, usually diamond-like-carbon (DLC) to protect against occasional impacts that might occur and against the corrosion of the magnetic layers beneath the DLC. A molecularly thin film of lubricant (typically about 1 nm thick) is applied over the DLC to further ensure that the head disk interface (HDI) is not damaged by prolonged contact or occasional high-speed impacts. With the demand of increased areal density, resulting in reduced spacing, there is an increased probability of high-speed contacts between the slider and the disk. Concurrently, the protective overcoat thickness needs to be reduced to ensure smaller magnetic spacing.

Prolonged or occasional head/disk contacts not only can damage the disk due to large dynamic contact forces (Katta et al. 2009a) but also can cause large vibrations of the recording slider in normal as well as lateral (off-track) directions. To improve the mechanical reliability of such ultra-low-flying HDIs, different approaches have been suggested, such as minimizing bouncing vibrations via ABS design (Knigge and Talke 2000), minimizing the contact force magnitude, making harder and denser overcoats (Guruz et al. 2001) and designing “smart” molecularly thin lubricants with preferential bonding and mobility (Mate et al. 1998). A major requirement to minimize the risk of head/disk contact is the use of molecularly smooth disks and sliders, and extremely low and stable flying sliders. Ultra-low clearance HDIs with such super-smooth media and sliders will result in strong attractive intermolecular adhesive forces that occur when two extremely smooth interfaces come into proximity or contact each other (Israelachvili 1991). Thus, adhesive forces should also be considered in the design of robust and reliable HDIs.

The reduction in the magnetic spacing cannot simply be achieved by improvements in the slider/disk interfaces since the elements involved such as lubricant, overcoat, roughness, and spacing are reaching their physical limit. Hence, alternative approaches have been advanced. One such successful development is a slider with “thermal

fly-height control” (TFC): A thin-film piezoelectric heater element is integrated into the recording head and is used to actively control the clearance to less than few nanometers. On supplying a voltage, the heater that carries the reader and writer elements, also called the thermal pole tip protrusion (T-PTP), protrudes closer to the disk. With this technology, the magnetic spacing is further reduced to less than 5 nm. This also means that there is an increased chance of contact of the T-PTP with the disk.

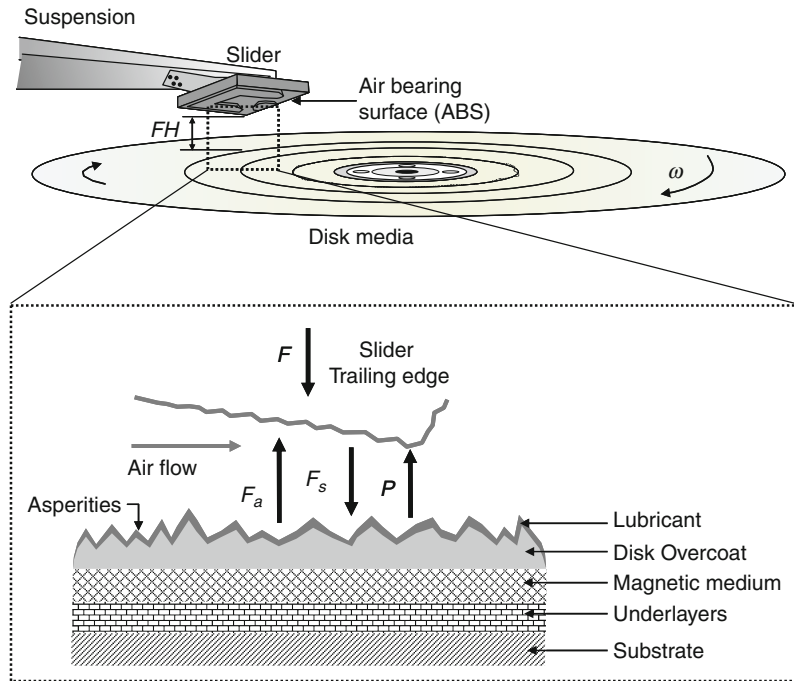
For further development of magnetic storage HDD technology, it is critical to understand and model the contact interactions in a HDI. Contact in HDI results in unwanted forces. These forces cause stresses and eventual damage to the HDI. Due to the presence of the extremely thin lubricant on the magnetic disk, a situation can also occur where the high forces (or pressures) will influence the lubricant to take up part of the applied forces. This phenomenon is called sub-boundary lubrication. Due to the complexities involved, conventional analytical modeling approaches are difficult to adopt to understand the behavior of the HDI. Approaches such as numerical simulations, atomistic/molecular dynamics simulations, and fast Fourier transform-based approaches have also been used to model contacting/sliding surfaces with some success.

HDI Contact Modeling

Head Disk Interface Issues

Light sporadic contact can occur during normal operation of a HDD. In such cases the effects of roughness and adhesive forces become important and might not necessarily result in head or disk damage, but may result in bouncing vibration and off-track vibration (motion) of the slider. Hence HDI contact cannot simply be viewed as a static phenomenon but is coupled with the dynamic behavior of the overall system. Such coupled contact/friction/dynamic interactions are termed microtribodynamics (Lee and Polycarpou 2005). Traditionally, HDI contacts were studied for a nominally flat slider interacting with the rotating disk. However, with the implementation of the T-PTP technology, the pole tip will contact the disk instead of the slider. The pole tip that protrudes closer to the disk is basically a permalloy material that expands on heating. Its geometry has to be accounted for to study such nanoscale contact behavior.

In summary, to successfully model and understand the contact interactions in a dynamic HDI, different elements need to be considered. These include precise contact geometry, material behavior/properties, surface topography (roughness), contact force (P), adhesion force (F_s),



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 1 Schematic of read/write recording head flying over a magnetic disk

air bearing force (F_a), friction force (Q), effect of lubricant, and the effect of these parameters on the dynamics of the HDI (Fig. 1).

Surface Topography

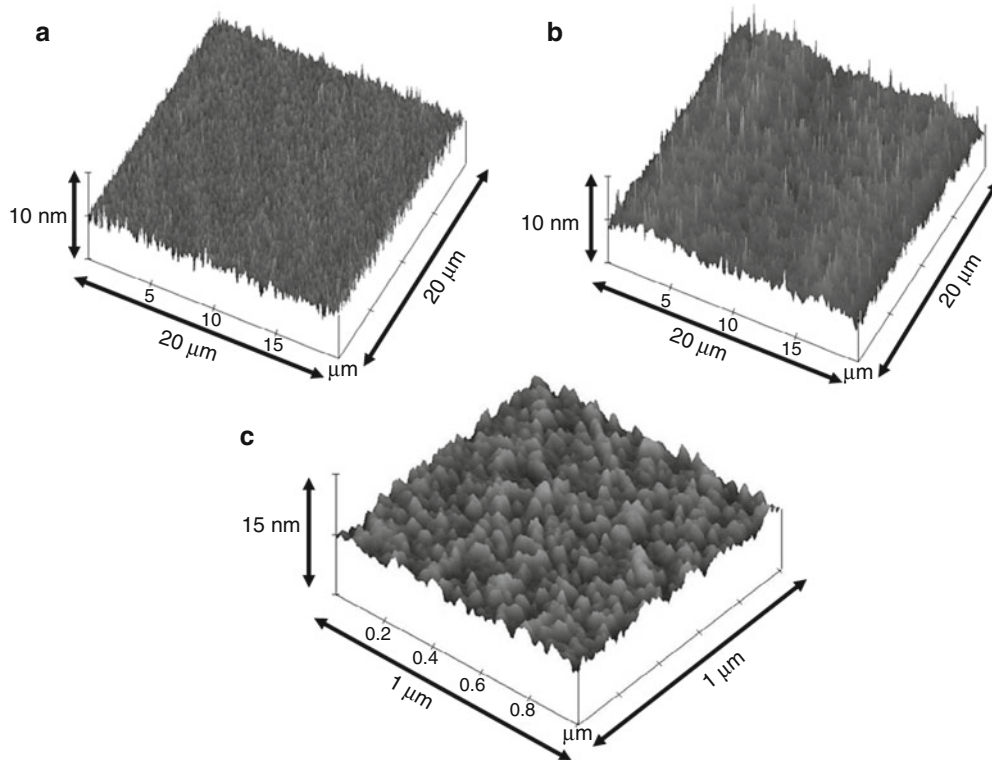
Surface topography or roughness is a very important factor in achieving the required magnetic spacing between the slider and the disk. Rough surfaces result in more frequent intermittent contact events between the slider and the disk. To avoid such events, current roughness values of the disks used in HDD are in the sub-nanometer range, thus super-smooth (or atomically smooth). Nevertheless, even in such cases the surfaces will not be mathematically flat (or infinitely smooth). Ideally, a perfectly smooth surface is needed for better HDD reliability. However, as it can be seen in high-resolution surface scans (of the magnetic disk and the slider) in Fig. 2 that the surfaces are actually rough at a pertinent scale. Moreover, making surfaces even smoother will result in an increased effect of adhesion at the HDI.

Certain roughness quantities, which are measures and descriptors of the surface profile, represent the extent of roughness on a surface. Roughness parameters that are commonly used to describe the surface are simple amplitude descriptors of center-line-average

(CLA) R_a and root-mean-square (RMS) σ (standard deviation of surface heights). From these two parameters, the most accepted measure is the σ value since it is a better statistical property. Amplitude descriptors, however, are influenced by the scale of measurement, such as the scan size. To reduce the scale dependency on these measurements, functional filtering of the rough surface data is recommended (Suh and Polycarpou 2006). As stated, this single numerical parameter provides information of the surface in the vertical direction only. To capture the effects of the surface, spatial and hybrid descriptors are also needed. Such spatial measures are obtained by considering that the rough surface is covered with a large number of asperities (representing the peaks of the surface) whose summits are spherical in shape with a constant radius of curvature R . Within a certain nominal surface area A_n , the number of these asperities N could be obtained. This descriptor is called the asperity areal density η ($= N/A_n$). The roughness parameters corresponding to Fig. 2 are shown in Table 1.

HDI Contact

When two nominally flat surfaces come in contact, surface roughness causes contact to occur at discrete contact points. The contact load is determined by the number of



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 2 Typical AFM roughness images: (a) magnetic disk, (b) slider surface, (c) zoomed-in view of magnetic disk

Microtribodynamics of Magnetic Storage Hard Disk Drives, Table 1 Measured surface roughness parameters

Individual roughness parameters			
		Slider	Disk
σ	nm	0.64	0.38
R	μm	7.15	7.63
η	μm^{-2}	8.85	9.94

contact points, which depends on the surface topography. Greenwood and Williamson (1966) introduced a comprehensive theory to model elastic rough surface contact by assuming that the rough surface is covered with a large number of asperities whose summits are spherical in shape with the same radius of curvature. Moreover, two rough surfaces in contact can be replaced with an equivalent rough surface in contact with a smooth flat as shown in Fig. 3. The equivalent roughness parameters for the slider and disk in Table 1 are shown in Table 2. The asperity heights above a reference plane (mean of asperity heights)

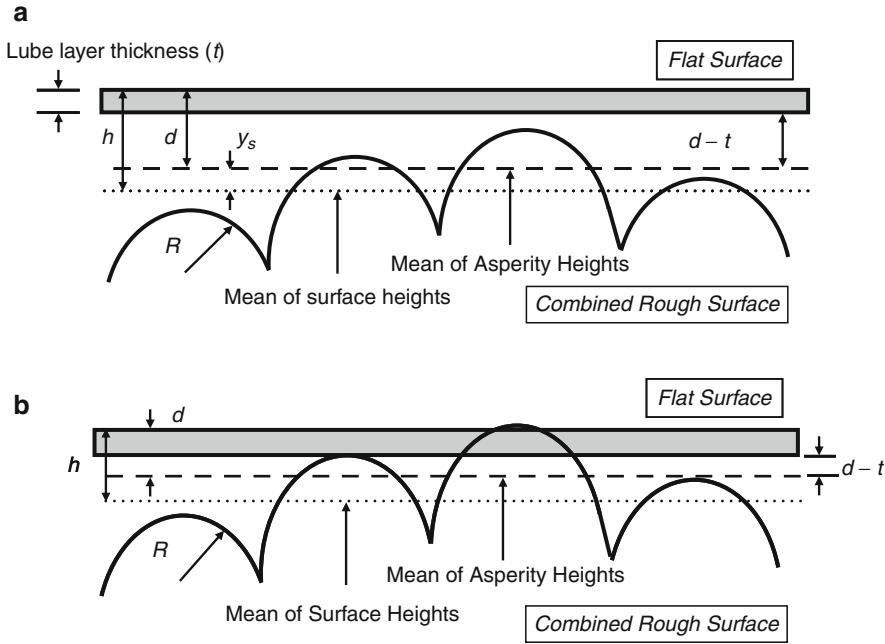
follow a certain distribution such as Gaussian and are represented by a normal probability distribution function

$$\phi(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (1)$$

where z is the asperity height and σ is the standard deviation of the asperity heights.

The elastic contact behavior of each individual asperity can be obtained using Hertzian contact equations. The contact behavior of an asperity when it experiences plastic deformation has also been studied and proposed models can be found for example in (Chang et al. 1987). When the two surfaces approach each other until their reference planes are separated by a distance d , then all the asperities whose height is greater than d will make contact. The total load and the real contact area can then be obtained by statistical summation of the force and contact area of all asperities in contact. The probability of making contact at any asperity of height z is given by

$$\text{prob}(z > d) = \int_d^{\infty} \phi(z) dz \quad (2)$$



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 3 Equivalent isotropic rough surface in contact with an infinitely smooth surface: **(a)** before the onset of contact, $h > 3\sigma$, **(b)** in contact, $h < 3\sigma$

The total load due to contact of all contacting asperities between two surfaces is obtained by statistically summing the force on individual asperities obtained from Hertzian contact equations (or other advanced asperity contact equations). Based on Hertz contact, the total load is

$$P(d) = \frac{4}{3} N E_r R^{\frac{1}{2}} \int_d^{\infty} (z - d)^{\frac{3}{2}} \phi(z) dz \quad (3)$$

which is obtained from the Hertz (single asperity) elastic contact equation

$$P_i = \frac{4}{3} E_r R^{\frac{1}{2}} (z - d)^{\frac{3}{2}} \quad (4)$$

where E_r is the combined or reduced elastic modulus of the two contacting surfaces.

It has been observed that if a surface is repeatedly magnified, increasing details of roughness are observed down to nanoscales. In addition, the roughness at all magnifications appears similar in structure. Such behavior can be characterized by fractal geometry. Thus, contact between rough surfaces can be characterized alternatively using (theoretically) scale-independent fractal roughness parameters (Majumdar and Bhushan 1991; Yan and Komvopoulos 1998). A practical issue in using fractal

theory to model rough surfaces is that the fractal parameters that enter may not have a clear physical meaning and that realistic engineering surfaces do not truly possess scale independence.

Lubricant

In magnetic storage disks, a lubricant layer is also used to reduce corrosion as well as contact damage. In current disks, it is of the order of a nanometer thick or less, which means only a monolayer or sub-monolayer of lubricant is present on the disk surface and is designed to strongly adhere to the disk (at least some of the lubricant). For a strongly adherent molecularly thin lubricant, meniscus-based stiction will be negligible. Such form of lubrication is termed sub-boundary lubrication and there will be an adhesion force between the lubricant surface and the contacting slider surface. Hence in HDI contact models, the effect of the molecularly thin lubricant has to also be considered. The influence of the strongly adherent molecularly thin lubricant is such that it will affect the energy of adhesion $\Delta\gamma$ and consequently the adhesion force F_s . In this model, the contact force P and friction force Q remain unaffected by the lubricant since the lubricant is too “soft” to provide significant resistance during contact.

Microtribodynamics of Magnetic Storage Hard Disk

Drives, Table 2 Equivalent surface roughness parameters for slider and disk in Table 1

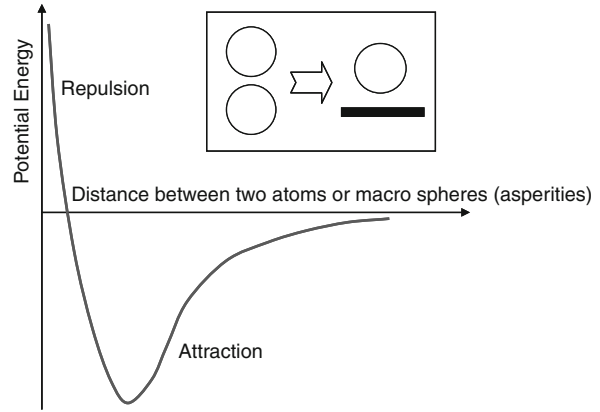
Equivalent roughness parameters		
		Slider/Disk
σ	nm	0.74
R	μm	5.22
η	μm^{-2}	9.33

Adhesion Force

As stated above, extremely smooth disk and slider surfaces are used in HDDs to allow nanometer spacing between the slider and disk and reduce the risk of contact. However, when two extremely smooth surfaces come into proximity and/or contact, strong intermolecular adhesive forces are present.

To understand and model these adhesive interactions (that occur when two continuum surfaces approach each other), consider a similar phenomenon that occurs as two atoms are brought close together and their electrons and nuclei interact, giving rise to intermolecular forces, as depicted schematically in Fig. 4. These atomic interactions can be modeled using an atomic potential energy such as the Lennard-Jones (L-J) potential, which is the most commonly used potential, and is mildly attractive (6th power of distance) as atoms approach each other from a distance, but strongly repulsive (12th power of distance) when they come too close to each other. This atomistic modeling was successfully extended to represent the intermolecular forces in continuum surfaces, and such approaches are known as mechanics-based adhesion models. Two commonly used models are the Derjaguin-Muller-Toporov (DMT) (Derjaguin et al. 1975) and the Johnson-Kendall-Roberts (JKR) (Johnson et al. 1971) adhesion models that were derived for single asperity contacts. One of the basic differences between these two models is that the DMT model assumes that the adhesion force originates from outside the contact region, while the JKR model assumes that the adhesion force exists only inside the region of contact. The atomic potential energy versus normal separation shown in Fig. 4 (for two atoms) also applies to single asperity contacts (continuum surfaces) using different parameters. Based on the L-J surface potential, the surface pressure $p(Z)$ due to adhesion is given by

$$p(Z) = \frac{8\Delta\gamma}{3\varepsilon} \left[\left(\frac{\varepsilon}{Z} \right)^3 - \left(\frac{\varepsilon}{Z} \right)^9 \right] \quad (5)$$



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 4 Schematic of Lennard-Jones interatomic or surface potential between two atoms or macro spheres (asperities), respectively

where ε is the equilibrium spacing, which is the minimum possible distance for contact to occur; $\Delta\gamma$ is the energy of adhesion, which is obtained by the summation of the surface free energies of the two surfaces; and Z is the separation of the two surfaces outside the contact area. The adhesive force $F_s(Z)$ between a sphere and a flat can be obtained by integrating $p(Z)$ over the spherical surface outside the contact region as follows:

$$F_s(Z) = 2\pi \int_a^\infty p(Z) r dr \quad (6)$$

where r is the radial distance measured from the center of the contact area, a is the radius of the contact given by $a = \sqrt{\omega R}$, and $\omega = z - d$, d is the separation of mean planes of asperity heights.

Engineering surfaces, including extremely smooth magnetic storage surfaces will still have some roughness (see Fig. 2), which can be represented using asperities whose heights follow a certain Gaussian distribution (see Fig. 3). Thus, a rough surface contact model like the Greenwood and Williamson model can be used to obtain the total adhesion force, which is the statistical summation of the individual asperity contributions. At any instance, these contributions can be grouped into three categories: (a) completely noncontacting asperities, (b) solid noncontacting-lubricant contacting asperities, and (c) solid contacting asperities (elastic, elastic-plastic, or fully plastic) (Chang et al. 1988a).

For a non-contacting asperity, integrate (5) for the profile of a whole sphere with radius R

$$F_s = \frac{8}{3}\pi R\Delta\gamma \left[\left(\frac{\varepsilon}{Z} \right)^2 - 0.25 \left(\frac{\varepsilon}{Z} \right)^8 \right] \quad (7)$$

For a solid noncontacting-lubricant contacting asperity, the adhesion force is obtained by setting $Z = \varepsilon$ in (5)

$$F_s = 2\pi R\Delta\gamma \quad (8)$$

For solid contacting asperities, the profile of the deformed asperity is needed to calculate the surface adhesive pressure. Assuming, the contact is only elastic, then using Hertz theory the profile of the deformed sphere will be

$$Z(r, a) = \frac{1}{\pi R} \left[a(r^2 - a^2)^{\frac{1}{2}} - (2a^2 - r^2) \tan^{-1} \left[\frac{r^2}{a^2} - 1 \right]^{\frac{1}{2}} \right] + \varepsilon; \quad r \geq a \quad (9)$$

where $Z(r)$ is the separation between asperity and the flat surface outside the contact region. Thus, the adhesion force can be obtained as

$$F_s = 2\pi \int_{rt}^{\infty} p(Z - t) r dr \quad (10)$$

where t is the lubricant thickness and rt is obtained from (9) and using $Z(rt, a) = t$.

Further advancements of these contact models have been performed where the exact profile of the deformed asperity is obtained when plastic deformation also exists; see, for example, (Kogut and Etsion 2002). Advanced models also account for the adhesion force on the inside as well as outside the contact area (Shi and Polycarpou 2005). Thus, the total contribution of the adhesion force during rough surface contact is obtained by the statistical summation of the individual asperity contributions where appropriate equations (7–10) are integrated depending on the regime a certain asperity is in.

Electrostatic forces F_{el} typically exist between surfaces that contain a certain amount of fixed electric charge that causes an electric field across the interface. Two main types of attractive electrostatic forces are the electrostatic image force based on classical Coulombic attraction, and the electrical double layer force based on electrostatic contact potential. The electrostatic image force can be important in HDIs, especially when charged contamination particles are trapped in the air-bearing surface (ABS). Charged particles contribute to the total electrostatic force due to image force, Coulomb force, and polarization acting on the particles. The electrostatic image force is also particle size dependent, and depending on the sign of charge,

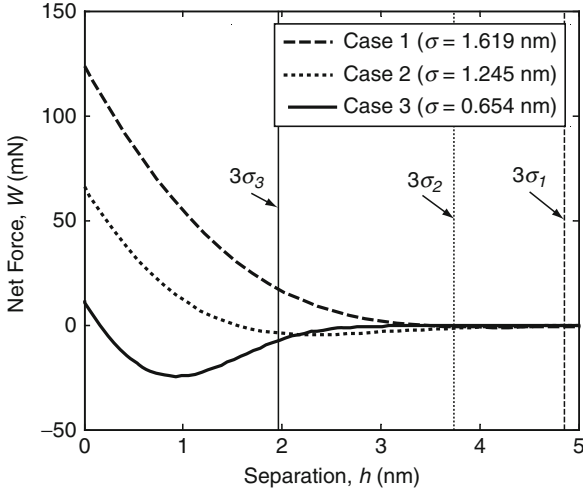
charged particles may attach themselves to either the disk or slider surface, thereby changing the slider trajectory and ABS performance. In micro/nanosystems composed of multilayered materials, such as the HDI, the electrical double layer force dominates over the electrostatic image force. The electrical double layer force, commonly referred to as electrostatic force, occurs due to an interfacial contact potential U caused by two dissimilar types of materials that are in contact or separated by a small distance. Different contact potentials at the HDI arise due to two different conducting parts of the slider and disk, namely, $\text{Al}_2\text{O}_3\text{--TiC}$ in the slider body and CoCr alloy magnetic layer in the disk body. Thus, the total potential difference at the HDI is the sum of an externally applied bias voltage and an internal contact potential. Generally, it is considered that the electrostatic attraction comes from the charges on surface layers of only a few angstroms, while the effect from those beyond the surface layers is neglected. Moreover, the electrostatic repulsion between two surfaces is considered negligible compared to the electrostatic attraction. The expressions for the electrical double layer force or electrostatic force for a sphere on flat is given by

$$F_e = (\pi \varepsilon_0 U^2 R) \left(\frac{1}{Z} \right) \quad (11)$$

where ε_0 is the permittivity of the medium, and U is the potential difference of the two material surfaces. For rough surface contact, the electrostatic force for a single asperity (11) can be statistically summed to obtain the total electrostatic force on the surface. This force will effectively increase the contribution of the adhesive force.

Net Force and Roughness in HDIs

To investigate the importance of roughness and its influence on the adhesive forces generated in a HDI, net force ($W = P - F_s$) calculations for three different measured roughness levels of HDI interfaces that represent a rough surface (case 1), smooth surface (case 2), and super-smooth surface (case 3) are depicted in Fig. 5. The adhesive force calculations are based on the ISBL model (Suh and Polycarpou 2005), including the effect of electrostatic force. When the net adhesive force is positive, the contact force is higher than the adhesive force, while the opposite is true when the net adhesive force is negative. The pull-off force is defined as the maximum negative net adhesive force, which is equivalent to the external input force required to separate two surfaces. The general trends delineated in Fig. 5 suggest that contact dominates the rough interface (case 1) as well as the intermediate interface (case 2), whereas the



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 5 Net force W vs separation h during contact between disk and slider surfaces for $A_n = 1,000 \mu\text{m}^2$ showing the onset of contact separation at 3σ

adhesion dominates the extremely smooth interface (case 3). The magnitudes of the corresponding pull-off forces (maximum negative external force) are 1.1 mN for case 1, 4.6 mN for case 2, and 24.4 mN for case 3.

Friction

When the slider makes contact with the spinning disk, a lateral friction force also acts on the slider, which can render the slider unstable as well as cause wear. Hence, it is also important to accurately model the friction at the HDI. The classical definition of static friction coefficient μ is that it is the ratio of the tangential shear force Q needed to shear the junctions between the contacting asperities and the external normal load W (perpendicular to the plane of contact). The actual contact load P in the true area of contact differs by the adhesion force F_s ,

$$\mu = \frac{Q}{W} = \frac{Q}{P - F_s} \quad (12)$$

The contact force P can be obtained from Hertzian or other elastic-plastic contact equations and the adhesion force F_s can be obtained as discussed in above sections. To estimate the static friction coefficient during contact, the tangential shear force Q must be calculated, including rough surface contact. It is assumed that plastically deformed asperities cannot support any tangential load. Only the contacting asperities that have not yet reached their elastic limit can contribute significantly to resisting sliding. During rough surface contact, on the application

of normal load, some asperities are stressed more than others. As the tangential force is increased, the contacting asperities reach a point of failure, beyond which relative sliding of the surfaces occurs. Static friction force is defined as the sum of all the tangential forces of all the individual pre-stressed asperities. For this purpose the maximum shear force Q_i is required for a single asperity junction to be sheared for a given applied contact load P_i . Analytical expressions were derived by Chang et al. (1988b) with the following functional form:

$$Q_i = f\left(\frac{\omega}{\omega_c}, v\right) P_i \quad (13)$$

$$\omega_c = \left(\frac{\pi KH}{2E}\right)^2 R \quad (14)$$

where interference $\omega = z - d$, $K = 0.454 + 0.41v$, v is Poisson's ratio, H is hardness, and ω_c is critical interference beyond which failure (plastic deformation) occurs in the asperity. The specific form of the function f in (13) depends on whether the failure occurs on or below the surface. Consequently the total tangential shear force for rough surface contact can be obtained by combining (3) and (13) and recalling that only elastically deforming asperities contribute

$$Q(d) = \frac{4}{3} N E R^{\frac{1}{2}} \int_d^{d+\omega_c} (z-d)^{\frac{1}{2}} f\left(\frac{z-d}{\omega_c}, v\right) \phi(z) dz \quad (15)$$

Thus, to successfully model these interactions, the following parameters shown in Table 3 are required. Note that the following list is only indicative of the kind of parameters needed to model contact in magnetic storage disk drives described here and not a comprehensive list.

HDI Dynamic Contact Modeling

HDI Dynamic Model

Under small HDI spacing of less than 5 nm, it is critical to model HDI interactions, which involve surface topography, adhesion, contact, and friction forces, coupled with HDI dynamics. The study of such interfacial and dynamic coupling is termed microtribodynamics. In earlier studies it was shown that with sub-nanometer slider and disk RMS roughness required for high recording densities, strong intermolecular adhesive forces are significant and can lead to premature HDI failures. For example, in a fully flying HDI, the slider is either pulled down towards the high-speed rotating disk causing a "head crash," or

Microtribodynamics of Magnetic Storage Hard Disk Drives, Table 3 List of parameters required to model the contact interactions in magnetic storage HDI

Symbol	Parameter
σ	RMS value of asperity heights; also standard deviation of asperity heights
R	Asperity radius of curvature
η	Number of asperities present per unit area; asperity density
A_n	Nominal contact area
ν_{slider}, ν_{disk}	Poisson's ratios of the slider and disk
E_{slider}, E_{disk}	Elastic moduli of the slider and disk
$E_r = \left[\frac{1-\nu_{slider}^2}{E_{slider}} + \frac{1-\nu_{disk}^2}{E_{disk}} \right]^{-1}$	Combined or reduced elastic modulus
ε	Equilibrium spacing between two surfaces
$\gamma_{slider}, \gamma_{disk}$	Surface free energies of slider and disk
$\Delta\gamma = \gamma_{slider} + \gamma_{disk} - \gamma_{slider}\gamma_{disk}$	Energy of adhesion
U	Electrostatic potential between slider and disk
t	Lubricant thickness

exhibits large flying-height modulation (FHM). It is therefore necessary to develop models that include these coupled interactions. One such model is detailed here and the model uses measured roughness, material and geometrical parameters and does not include any fitting parameters or any assumed friction coefficient values.

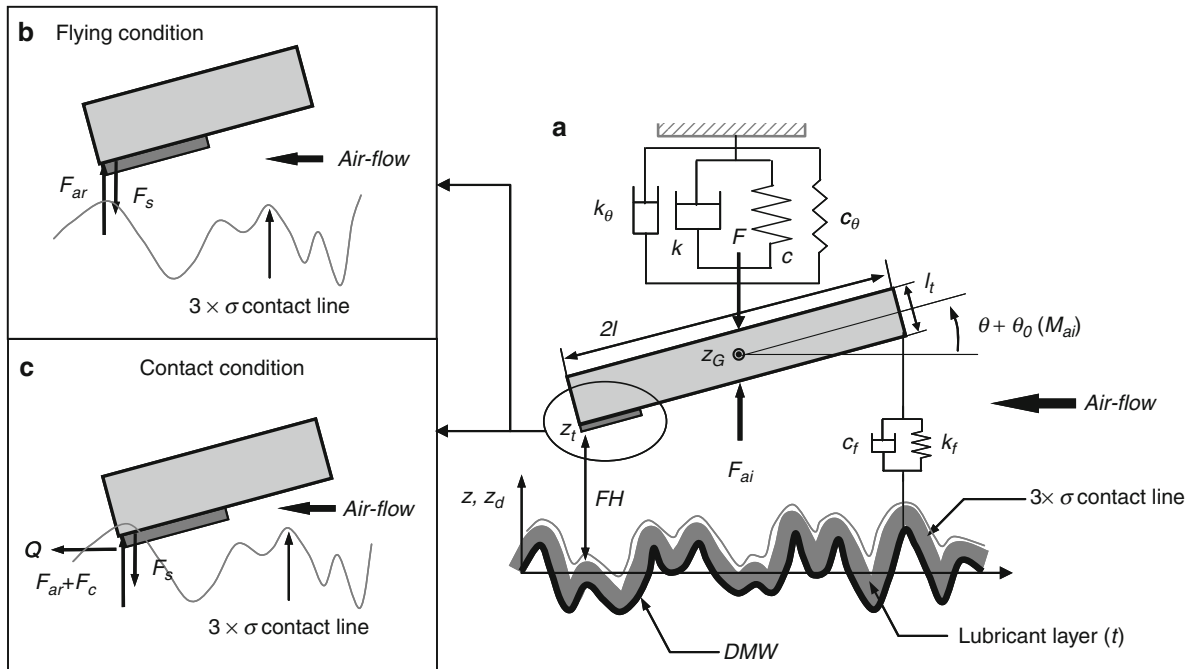
Figure 6 depicts a typical 2-DOF dynamic contact model of the HDI (Lee and Polycarpou 2005). The tri-pad slider used in these simulations has length $2l$ and thickness l_r . The soft suspension system is represented by a constant pre-load F applied to the slider, and linear translational and rotational springs k and k_θ , and damping c and c_θ elements, respectively. The model also includes adhesion and roughness effects, as shown in Fig. 6a. When the slider is flying without contacting the disk, the rear air-bearing force, F_{ar} and linear damping c_r are applied to the trailing edge (TE) of the slider and are away from the center of the mass by l (where normal slider displacement is z_G). While F_{ar} is nonlinear and increasing sharply with decreasing FH, a second linear air-bearing force term represented by a linear spring k_β and damping c_β at the leading edge (LE) of the slider is used.

The adhesive force, F_s , is applied at the TE of the slider and this nonlinear force also increases sharply with decreasing FH. Depicted in Fig. 6b is the preferred equilibrium state where the slider dynamics are dependent on air-bearing, adhesion, and suspension dynamics. When the slider makes contact with the disk (Fig. 6c), the interfacial contact force F_c (including linear contact damping c_c) and friction force Q (in addition to F_{ar} and F_s) are also acting at the TE of the slider due to the pitch motion (θ). Also, note that the onset of contact is based on a 3σ ($3 \times$ RMS roughness) contact condition.

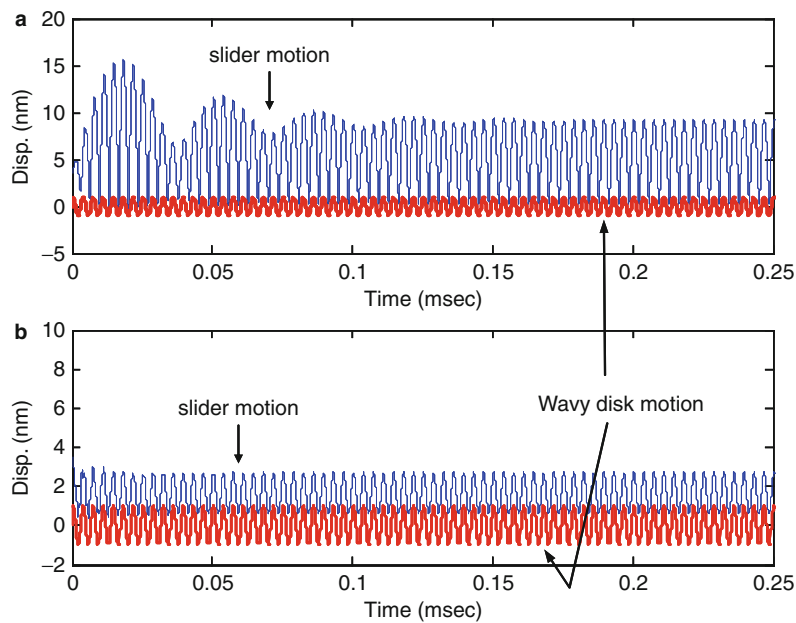
A harmonic excitation, which could come from the wavy motion of the disk (and possibly structural dynamics), termed dynamic microwaviness (DMW) while the disk spinning, could be used as an input for simplicity (280 kHz in this case) and the model can be simulated numerically to understand the dynamic behavior of the HDI. Figure 7 shows typical simulated results without considering the adhesive force (Fig. 7a) and adhesion included (Fig. 7b). When the adhesive force was not considered, even though the slider was flying at its nominal FH of 3 nm, it was exhibiting significantly large bouncing vibrations. After the initial transient response, the slider is bouncing off the wavy disk surface from 0.78 to 7.7 nm. When the adhesive force was considered at the HDI, the slider was not able to fly and was periodically contacting the wavy disk surface after a short initial transient period (see Fig. 7b). The slider motion of Fig. 7b exhibits very small bouncing vibration from 1.10 to 1.22 nm. Since the maximum bouncing vibration is less than 3σ , i.e., 1.72 nm (contact condition), the slider is physically dragging over the wavy disk motion, i.e., “perfect tracking.”

HDI Dynamics with Thermal Protrusion

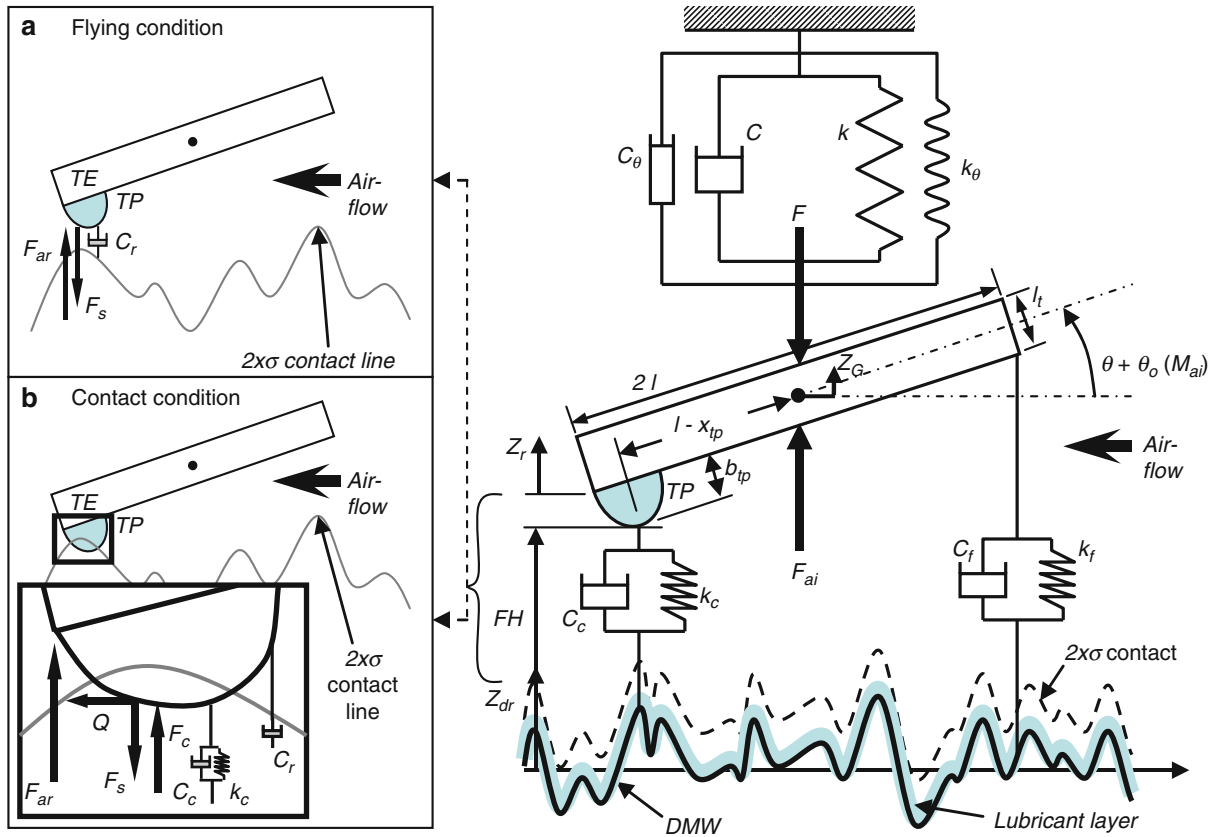
Current advances in the technology of hard disk drives (HDDs) have established the use of thermally actuated pole tip protrusion (T-PTP), herein referred to simply as thermal protrusion (TP), as an effective method to practically reduce the head-media spacing (HMS) to nanometer scale to achieve ultra-high density recording. The principle behind the operation of such a system is to fly the slider at a nominal flying height (FH), where the air-bearing dynamics are steady and predictable, and use a protrusion that is situated at the trailing edge (TE) of the slider to bring the read-write elements closer to the disk. This protrusion can translate the read-write elements several nanometers towards the disk surface and, through the use of a feedback system, its height can be regulated accordingly to allow for “non-contacting” tracing of the disk motion while minimizing the HMS and flying height modulation (FHM).



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 6 Schematic of 2-DOF dynamic HDI contact model. **(a)** Overall model along with the suspension system, **(b)** model state during flying condition (non-contact), **(c)** model state when contact occurs



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 7 Typical dynamic simulation results from 2 DOF dynamic HDI contact model when (a) adhesive forces are not considered, (b) adhesive forces are considered (280 KHz harmonic input)



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 8 Schematic of 2-DOF dynamic HDI contact model with thermal protrusion: (a) shows the flying condition where only air-bearing and adhesion forces are acting, (b) shows the contact condition where contact and friction forces also come into play

Extending the above dynamic HDI contact model, the HDI dynamic models should include the geometry of the TP since this is the component of the slider that first contacts the disk. A typical model, which is similar to the 2-DOF model described in the earlier section, where the TP is also modeled is shown in Fig. 8 (Vakis et al. 2009). The TP is modeled in such a way that its height in the direction perpendicular to the bottom surface of the slider changes according to the power (current) applied to it to protrude towards the disk. One goal of such dynamic modeling is to identify the optimum TP height just before the interacting forces will lead to jump-to-contact condition. Additionally, DMW of a spinning magnetic disk will not be a steady-state excitation. It can be measured experimentally and the excitation can be a direct input to the HDI dynamic model, as has been done in this case.

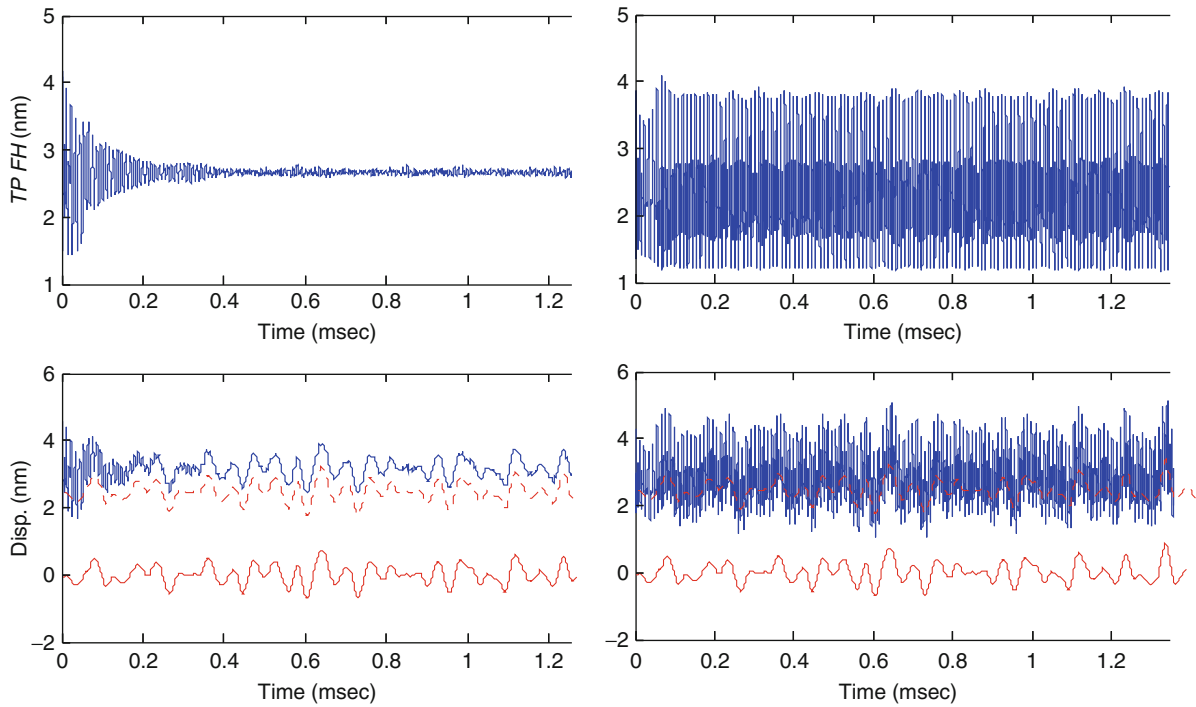
Figure 9 shows a plot of the thermal protrusion flying height time history for the cases when the TP height (size) increases from 6.7 to 6.8 nm. It can be clearly observed that contact is initiated for a TP height value somewhere in

between these two values. The TP flying height versus TP height plot (using the 3σ criterion), as shown in Fig. 10, clearly shows the initiation of contact for a TP height of 6.8 nm. In Fig. 10, the 3σ contact criterion is defined as whenever the TP flying height is less than $3\sigma - T_0$ ($T_0 = t$, lub thickness).

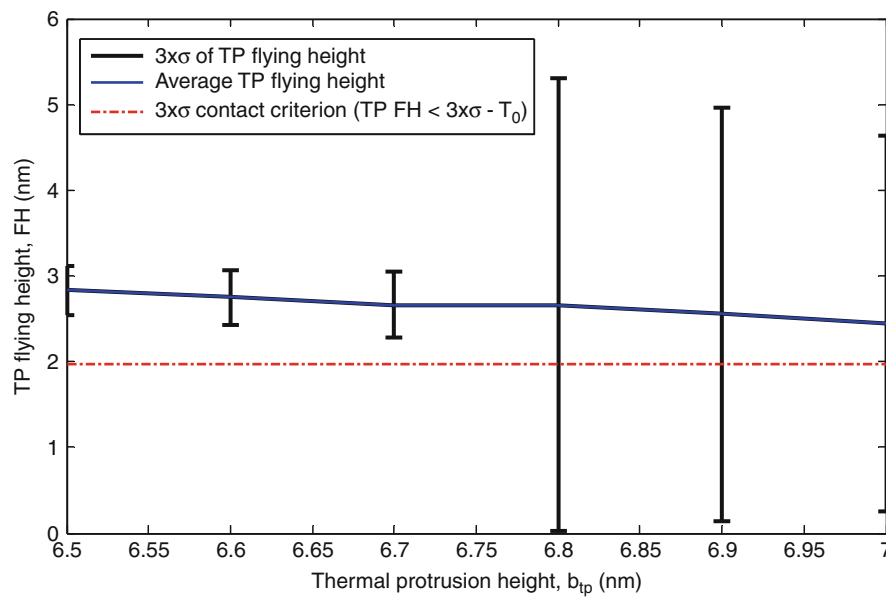
Further Contact-Related HDI Issues

Head Disk Scratch

There are situations when the head or particles caught between the slider and disk can scratch the disk and result in permanent damage on the disk and magnetic data erasures (Furukawa et al. 2008). Analytical solutions do not exist to estimate this damage due to the complexities associated with modeling the multiple thin film layers on a disk. Hence, numerical simulations using the finite element method are employed to model and examine the disk damage, for example see (Katta et al. 2010).



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 9 Dynamic simulation results showing jump-to-contact: plot of the TP flying height time history compared to the disk dynamic profile DMW using the 3σ contact criterion for TP height of 6.7 nm (left) and 6.8 nm (right)



Microtribodynamics of Magnetic Storage Hard Disk Drives, Fig. 10 Dynamic simulation results: plot of TP flying height against TP height showing initiation of contact using the 3σ contact criterion

Head Disk Impact

Due to the increased use of hard disk drives in mobile computing applications, occasional head disk impact might also occur when those mobile applications are subject to shock. When the head impacts the rotating disk, the high contact pressures as well as the high temperatures encountered can lead to loss of data or even catastrophic damage. The flash temperatures can be calculated during such impacts to examine whether thermally induced magnetic erasures occur (Yu et al. 2008). Similarly, elastic and elastic-plastic contact mechanics theories can be used to estimate the mechanical damage on the disk due to these impacts (Katta et al. 2009b).

Cross-References

- [ABS Designs](#)
- [Adhesive Contact of Elastic Bodies](#)
- [Asperities](#)
- [Contacts Considering Adhesion](#)
- [Contact Elasto-Plasticity](#)
- [Contact of Layered Materials](#)
- [Surface Roughness](#)

References

- W.R. Chang, I. Etsion, D.B. Bogy, An elastic-plastic model for the contact of rough surfaces. *J. Tribol.* **109**(2), 257–263 (1987)
- W.R. Chang, I. Etsion, D.B. Bogy, Adhesion model for metallic rough surfaces. *J. Tribol.* **110**(1), 50–56 (1988a)
- W.R. Chang, I. Etsion, D.B. Bogy, Static friction coefficient model for metallic rough surfaces. *J. Tribol. (Trans. ASME)* **110**(1), 57–63 (1988b)
- B.V. Derjaguin, V.M. Muller, Y.P.J. Toporov, Effect of contact deformations on the adhesion of particles, *Coll. Interface Sci.* **53**, 314–326 (1975)
- M. Furukawa, J.G. Xu, Y. Shimizu, Y. Kato, Scratch-induced demagnetization of perpendicular magnetic disk. *IEEE Trans. Magn.* **44**(11), 3633–6 (2008)
- J.A. Greenwood, J.B.P. Williamson, Contact of nominally flat surfaces. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **295**(1442), 300–319 (1966)
- M.U. Guruz, V.P. Dravid, Y.W. Chung, M.M. Lacerda, C.S. Bhatia, Y.H. Yu, S.-C. Lee, Corrosion performance of ultrathin carbon nitride overcoats synthesized by magnetron sputtering. *Thin Solid Films* **381**(1), 6–9 (2001)
- J.N. Israelachvili, *Intermolecular and Surface Forces*, 3rd edn. (Elsevier, 2011), Burlington, MA, USA
- K.L. Johnson, K. Kendall, A.D. Roberts, Surface energy and the contact of elastic solids. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **324**(1558), 301–313 (1971)
- R.R. Katta, A.A. Polycarpou, J.V. Hanchi, R.M. Crone, High velocity oblique impact and coefficient of restitution for head disk interface operational shock. *J. Tribol.* **131**(2), 021903 (2009a). 9 pp
- R.R. Katta, A.A. Polycarpou, J.V. Hanchi, M. Roy, Analytical and experimental elastic-plastic impact analysis of a magnetic storage head-disk interface. *J. Tribol.* **131**(1), 011902 (2009b). 10 pp
- R.R. Katta, A.A. Polycarpou, S.-C. Lee, M. Suk, Experimental and FEA scratch of magnetic storage thin-film disks to correlate magnetic signal degradation with permanent deformation. *J. Tribol.* **132**, 021902 (2010). 11 pp
- B. Knigge, F.E. Talke, Contact force measurement using acoustic emission analysis and system identification methods. *Tribol. Int.* **33**(9), 639–646 (2000)
- L. Kogut, I. Etsion, Elastic-plastic contact analysis of a sphere and a rigid flat. *J. Appl. Mech. Trans. ASME* **69**(5), 657–662 (2002)
- S.-C. Lee, A.A. Polycarpou, Microtribodynamics of pseudo-contacting head-disk interfaces intended for 1Tbit/in². *IEEE Trans. Mag.* **41**(2), 812–818 (2005)
- A. Majumdar, B. Bhushan, Fractal model of elastic-plastic contact between rough surfaces. *J. Tribol.* **113**(1), 1–11 (1991)
- C.M. Mate, P.H. Kasai, G.W. Tyndall, C.H. Lee, V. Raman, D.J. Pocker, R.J. Waltman, Investigation of phosphazene additive for magnetic recording lubrication. *Magn. IEEE Trans.* **34**(4), 1744–1746 (1998)
- X. Shi, A.A. Polycarpou, Adhesive transition from noncontacting to contacting elastic spheres: extension of the Maugis-Dugdale model. *J. Colloid Interface Sci.* **281**(2), 449–457 (2005)
- A.Y. Suh, A.A. Polycarpou, Adhesive contact modeling for sub-5-nm ultralow flying magnetic storage head-disk interfaces including roughness effects. *J. Appl. Phys.* **97**, 104328 (2005)
- A.Y. Suh, A.A. Polycarpou, Digital filtering methodology used to reduce scale of measurement effects in roughness parameters for magnetic storage supersmooth hard disks. *Wear* **260**(4–5), 538–548 (2006)
- A.I. Vakis, S.-C. Lee, A.A. Polycarpou, Dynamic head-disk interface instabilities with friction for light contact (surfing) recording. *IEEE Trans. Mag.* **45**(11), 4966–4971 (2009)
- W. Yan, K. Komvopoulos, Contact analysis of elastic-plastic fractal surfaces. *J. Appl. Phys.* **84**(7), 3617–3624 (1998)
- N. Yu, A.A. Polycarpou, J.V. Hanchi, Elastic contact mechanics-based contact and flash temperature analysis of impact-induced head disk interface damage. *Microsyst. Technol.* **14**(2), 215–227 (2008)

Microtribology of Living Cells

- [Tribology of an Epithelial Cell Monolayer](#)

Mild Adhesion

- [Gear EHL Film Thickness and Wear Risk Analysis](#)

Milling

- [Gear Manufacturing Machines](#)

Mineral Oil

- [Mineral Oil Base Fluids](#)

Mineral Oil Base Fluids

BRIAN L. PAPKE

Shell Global Solutions (US) Inc, Houston, TX, USA

Synonyms

Base fluid; Base oil; Base stock; Mineral oil

Definitions

Mineral Oil: A liquid by-product or hydrocarbon fluid derived from the distillation of petroleum crude oil.

Base Stock: A lubricant component produced by a single manufacturer to the same specifications independent of feed source or manufacturer's location (API 1509 2005).

Base Oil: A single base stock or blend of base stocks used in a lubricant product.

Characteristics

Manufacture

The manufacture or refining of mineral oil base fluids from petroleum crude consists of up to five basic steps: (1) distillation (atmospheric and vacuum), (2) deasphalting, (3) solvent or hydrogen processing, (4) solvent or catalytic dewaxing, and (5) hydrogen or clay finishing (Sequeira 1994). These steps ensure that the finished product meets the performance and quality requirements of the finished lubricant formulation. The refiner will select the crude oil and utilize a combination of processes to economically refine mineral oils with the desired properties.

The distillation process separates hydrocarbons on the basis of volatility into viscosity cuts suitable for lubricant applications. A vacuum distillation is required to prevent thermal decomposition of the hydrocarbon molecules. Vacuum distillation is followed by an "upgrading" process to decrease the amount of aromatics and remove undesirable components such as sulfur and nitrogen-containing molecules. This process improves the oxidation resistance and increases the viscosity index. Dewaxing is required to remove the majority of paraffinic waxes that crystallize at low temperatures (at or above the pour point). Solvent dewaxing (a crystallization-filtration process) is being replaced by the more cost-efficient catalytic dewaxing process. In catalytic dewaxing, the molecular structures of the waxy paraffinic molecules are rearranged into more highly desirable branched paraffins to improve low temperature properties. A finishing process (hydrofinishing or



Mineral Oil Base Fluids, Fig. 1 Typical mineral oil base fluids

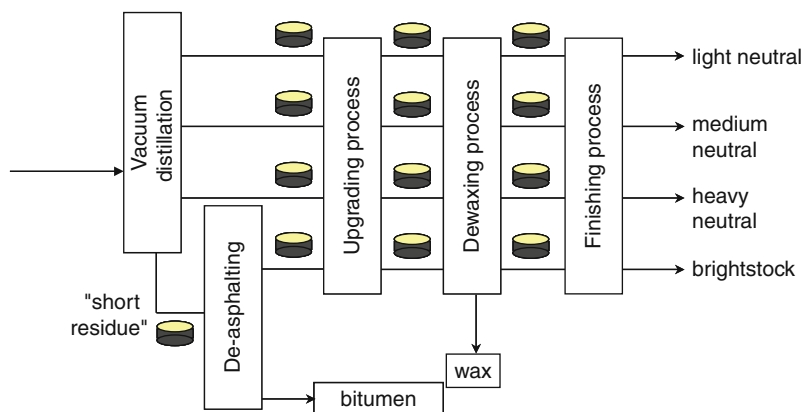
clay finishing) ensures the finished product meets color and oxidation stability standards when the residual sulfur content (a natural antioxidant) is very low. Base fluids exhibit a wide range of color variation depending on the feedstock used in manufacture and the refining process (Fig. 1).

A simplified schematic representative of typical mineral oil base fluid manufacture is shown in Fig. 2.

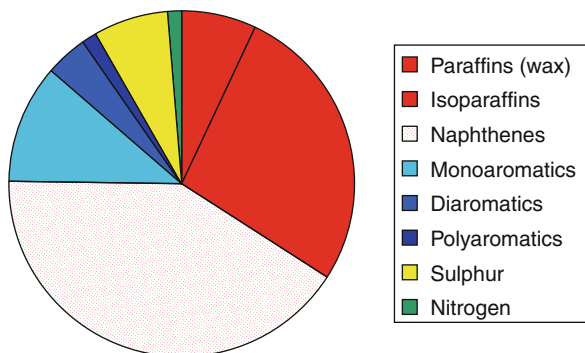
Composition

Mineral oil base fluids consist largely of carbon and hydrogen (hydrocarbons), but may also contain smaller amounts of sulfur, oxygen, and nitrogen compounds. Mineral oil fluids with viscosity properties suitable for lubricants have carbon numbers in the C₂₀–C₄₀ range. Because of the large number of different molecular structures possible in this carbon number range, compositions are described by major hydrocarbon classes. A simple acronym to aid in description of mineral oil compositions is the word PIANO (Paraffins, Isoparaffins, Aromatics, Naphthenes, and Olefins). Paraffins, isoparaffins, and naphthenes are saturated hydrocarbons; aromatics are the only unsaturated structures in mineral oil base fluids, as olefins are generally removed during the manufacturing process. The composition of a typical "Group I" mineral oil is shown in Fig. 3. (Mineral oil classification is described in the next section.)

A ball-and-stick model of a branched paraffinic hydrocarbon molecule is shown in Fig. 4, and an aromatic



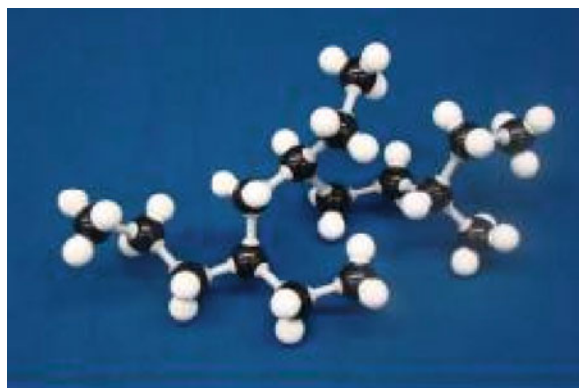
Mineral Oil Base Fluids, Fig. 2 Mineral oil manufacture process



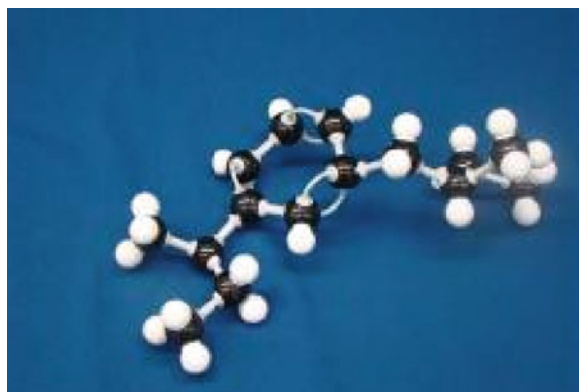
Mineral Oil Base Fluids, Fig. 3 Representative group-I mineral oil molecular composition

molecule representation is shown in Fig. 5. In both models, the length of the carbon-carbon single bonds has been extended to enable the “backbone” chain structure to be seen more clearly. Molecules are classified as “aromatic” if they contain unsaturated ring structures, however, much of the molecule is actually paraffinic. (In Fig. 5, note that mineral oil lubricants have higher carbon numbers than these models.)

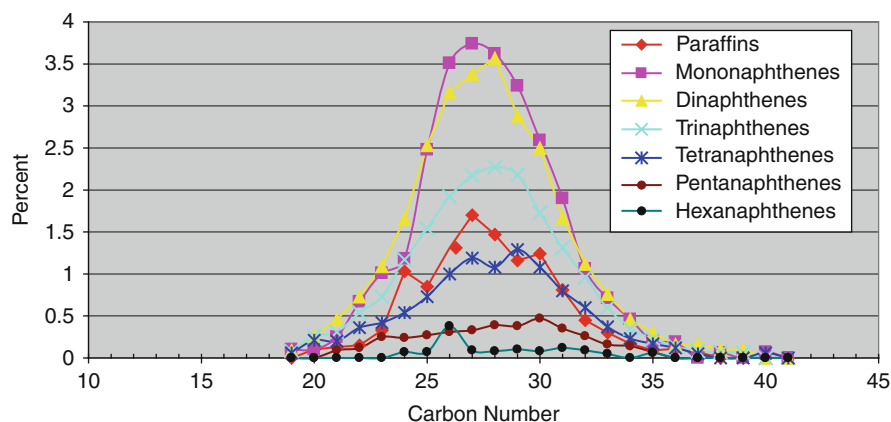
A more detailed analysis of the molecular composition of mineral base fluids can be obtained through analytical techniques such as field ionization mass spectroscopy (FIMS) (Altgelt and Boduszynski 1994). In this analysis, the oil is separated into aromatic and saturates fractions by column chromatography, and each fraction is then analyzed separately by FIMS. This technique generates parent ion mass spectra, and a hydrocarbon group-type analysis can be used to obtain the molecular distribution of major



Mineral Oil Base Fluids, Fig. 4 Branched paraffinic mineral oil molecular model



Mineral Oil Base Fluids, Fig. 5 Aromatic mineral oil molecular model



Mineral Oil Base Fluids, Fig. 6 FIMS analysis of the saturates fraction from a 6 cSt group-II base fluid

components present in each fraction. An example of a FIMS analysis of the saturates fraction from a 6 cSt (100°C) Group-II mineral oil is shown in Fig. 6. A FIMS analysis is particularly useful for the characterization of Gp II and Gp III base fluids, as it provides insight into molecular structure differences that can significantly impact base fluid performance. For example, polynaphthenic ring structures are known to be less oxidatively stable than paraffinic structures; the FIMS technique (Fig. 6) can be used to quantitatively compare base fluid molecular compositions and assess their relative oxidative stability.

Classification Schemes

Mineral oil base fluids are classified as either naphthenic or paraffinic as determined by crude source and defined by composition, viscosity index, and pour point. Naphthenic mineral oils have a high percentage of saturated cyclic carbon structures, and as a result have low viscosity indices and pour points relative to paraffinic mineral oils. Classification schemes developed jointly by the American Petroleum Institute (API) and the Technical Association of the European Lubricants Industry (ATIEL) place mineral oils into Groups I, II, III, or V based on composition (% saturates, % sulfur) and viscosity index (VI) (Table 1). Paraffinic mineral oils are classified in Groups I, II, and III; whereas, naphthenic mineral oils are classified in Group V.

Actual compositions within each group are narrower than those specified in Table 1. Typical compositions for Group I, II, and III mineral oils are shown in Table 2.

Group II and Group III mineral oil base fluids contain molecular structures formed during the manufacturing process that do not exist naturally in crude oils.

Group III fluids in particular are classified as synthetic fluids because of the high concentrations of manufactured molecular structures present.

Physical Properties

The physical properties of a mineral oil are critical to the performance of the finished lubricant product. The physical properties of the mineral oil provide a “baseline” performance; additive components blended into the formulation enhance critical performance aspects. Some of the key mineral oil properties include (a) viscosity, (b) oxidative stability, (c) volatility, (d) solvency, (e) pressure-viscosity coefficient, (f) pour point, and (g) color.

Viscosity is the single most critical mineral oil property. The change in viscosity with temperature is described by the viscosity index (Gresham 2007). Mineral oil viscosity at high temperatures and pressures may be determined by either the Barus or Roelands equations; viscosities of mineral oils increase exponentially with pressure (Fein 1998). At the high contact pressures experienced in cam-lifter contacts, mineral oil lubricants’ behavior may be described as more closely resembling a transient crystalline structure due to the highpressure viscosity coefficients of mineral oils.

A second critical property of mineral oils is the film thickness between moving parts. This property is a combination of viscometrics and pressure viscosity coefficient properties. Separation of moving parts is critical to wear protection, and a high pressure viscosity coefficient is beneficial in this regard. On the other hand, optimum fuel economy requires low viscosities and thin lubricant films – favored by low pressure viscosity coefficient values. Achieving an optimized mineral fluid mixture that

Mineral Oil Base Fluids, Table 1 Base stock classification scheme

	Group I	Group II	Group III	Group IV	Group V
Saturates	<90%	≥90%	≥90%		All basestocks not in Groups I, II, III, IV
	and/or	and	and		
Sulfur	>0.30%	≤0.03%	≤0.03%	PAOs	
	and	and	and		
VI	≥80 <120	≥80 <120	>120		

Mineral Oil Base Fluids, Table 2 Typical mineral oil compositions

Component	Base oil type		
	Group I	Group II	Group III
Saturates, %	65–85	93–99+	95–99+
Aromatics, %	15–35	<1–7	<1–5
Sulfur, ppm	300–3,000	5–300	0–30

provides adequate wear protection and simultaneous fuel economy benefits requires a careful base fluid selection. Group II and III base fluids have lower pressure viscosity coefficients than Group I and V base fluids, and are preferred blending stocks for many applications.

Key Applications

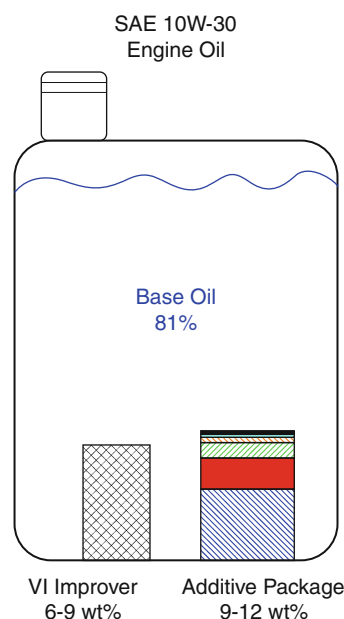
Scope

The material presented in this section is limited in scope to lubricant and grease applications for mineral oil base fluids. Mineral oils are also used in many non-tribological applications.

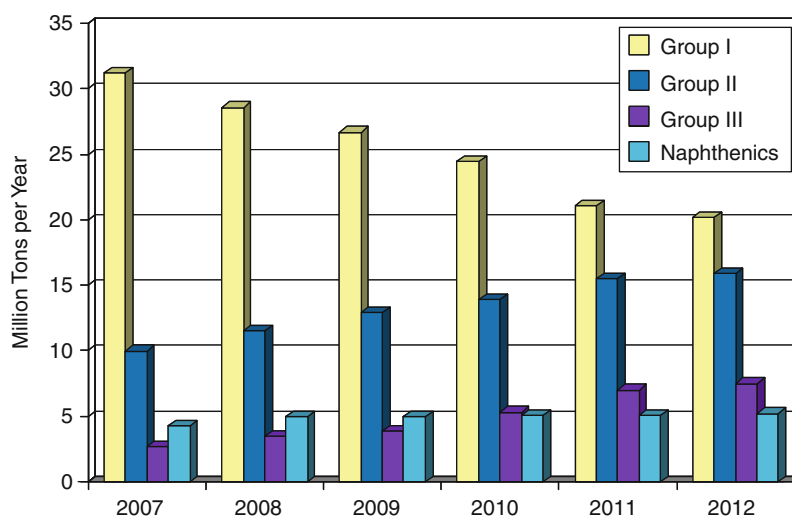
Supply and Demand

Mineral oils are the largest volume component of lubricants, typically accounting for more than 95% of the formulation. For example, whereas automotive engine oils may contain more than 80% of a mineral oil base fluid, additives added to the formulation (such as the viscosity index improver and the additive package) are prediluted in mineral oil (ca. 90% and 50%, respectively), so the actual base oil percentage in the formulation is 90–95% (Fig. 7). The physical properties of the mineral oil fluid therefore determine to a large extent the suitability of the finished product.

Global demand for mineral base fluids was approximately 46 million tons in 2007, and is expected to increase to 49 million tons by 2012 (Demarco 2008). In this relatively flat market, demand for Group I base stocks is

**Mineral Oil Base Fluids, Fig. 7** Composition of a typical automotive engine lubricant

expected to decline significantly, while demand for Group II/III fluids is expected to increase (Fig. 8) due to increasingly stringent performance requirements and environmental regulations.



Mineral Oil Base Fluids, Fig. 8 Global demand for mineral oil base fluids

Industrial Lubricants

Mineral oils are critical components in many industrial lubricants and greases. Viscosity and solvency are equally important properties for industrial lubricant products, particularly for greases, where the chemical nature of the base fluid affects both thickener structure and thickening power (Apparent viscosity of greases, 2001). Other industrial lubricant products that use mineral oil base fluids include turbine oils, gear oils, refrigerator oils, hydraulic oils, compressor lubricants, and shock absorber fluids.

Crankcase Lubricants

Major volume products for mineral oil base fluids include gasoline and diesel crankcase lubricants and driveline fluids. Applications include lubricants for gasoline and diesel engines, two-cycle engines, railroad diesel, marine diesel, and gas engine oils. These uses account for more than 60% of the demand for mineral oil base fluids. Crankcase lubricant products typically contain a greater additive percentage than industrial lubricants, as they must contend with a more severe operating environment.

Process Oils

While not strictly a tribological application, process oils represent approximately 10% of the global demand for mineral base oils. The chemical composition of the base stock is critical to their properties and applications. The function of process oils is application-dependent and includes areas such as (a) a component of a finished product (for example, an extender oil in rubber), (b) process aid such as defoamers or textile fiber lubricants, and

(c) physical or mechanical aids, such as gas absorption, or transformer oils.

Cross-References

- ▶ Aircraft Engine Lubricants
- ▶ Compressor Oils
- ▶ Effect of Lubrication on Fuel Economy in IC Engines
- ▶ Engine Condition Monitoring Based on Oil Analysis
- ▶ Engine Lubricants
- ▶ Gear Lubricants
- ▶ Locomotive Engine Oils
- ▶ Lubricant Viscosity
- ▶ Marine Lubricants
- ▶ Oil Life
- ▶ Paper Machine Lubricants
- ▶ Temperature and Pressure Dependence of Viscosity

References

- K.H. Altgelt, M.M. Boduszynski, *Composition and Analysis of Heavy Petroleum Fractions* (Marcel Dekker, New York, 1994)
- API 1509, Engine oil licensing and certification system, Appendix E. API Publishing Services, Washington, D.C. (2007)
- N.J. Demarco, External factors roll base oils, *Lubes'n' Greases*. LNG Publishing Company, Inc. **14**, 12–19 (2008)
- R.S. Fein, Liquid lubricants, in *ASM Handbook*, ed. by ASM International. Friction Lubrication and Wear Technology, vol. 18 (ASM International, Metals Park, 1998)
- R.M. Gresham, The basics of viscosity index. *Tribology and Lubrication Technology*, Sept 2007, 27–28
- B.L. Papke, R. Stamps, Apparent viscosity of greases. *Lubrication* **87**, 3 (2001)
- A. Sequeira Jr., *Lubricant Base Oil and Wax Processing* (Marcel Dekker, New York, 1994)

Minimum Quantity Cooling (MQC)

► [Tribology in Minimum Quantity Lubrication \(MQL\) Cutting](#)

Minimum Quantity Cooling Lubrication (MQCL)

► [Tribology in Minimum Quantity Lubrication \(MQL\) Cutting](#)

Minimum Quantity Lubrication (MQL)

► [Tribology in Minimum Quantity Lubrication \(MQL\) Cutting](#)

Minimum Quantity Lubrication (MQL) During Machining of Automotive Components

SHERI KURGIN
General Motors Powertrain Headquarters,
Pontiac, MI, USA

Synonyms

[NDR – nearly dry machining](#)

Definition

MQL is an environmentally friendly metal cutting fluid that utilizes a small amount of vegetable or synthetic oil combined with pressurized air to form a mist. MQL is an alternative to traditional wet machining, which usually consists of mineral or synthetic oil emulsified in a relatively large volume, measured in liters/min, of water. The primary motivations for MQL usage are to minimize the impact of manufacturing processes on the environment, improve working conditions, and reduce manufacturing costs (Stoll et al. 2007).

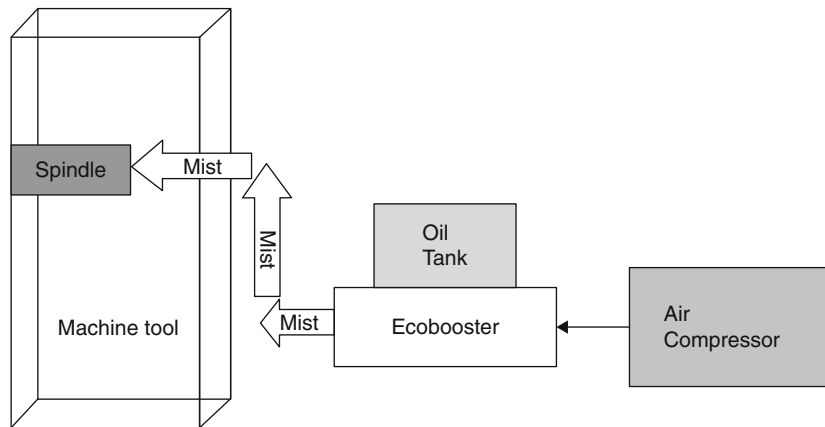
Scientific Fundamentals

Delivery Methods

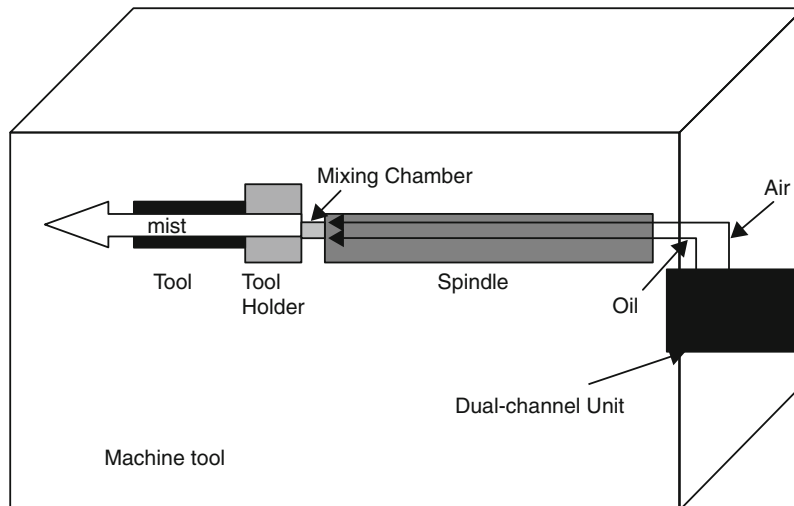
There are three basic MQL delivery methods. The first is external application. For external application, the compressed gas (usually air) and oil are mixed in a chamber and fed through a nozzle to the cutting zone. Studies have shown that both the direction of the nozzle and the distance from the cutting edge are significant. For example, distance of the nozzle from the cutting edge affects the uniformity of the droplet distribution; a uniform distribution of fine particles gives the most efficient lubrication (Wu and Chien 2007). The direction of the nozzle may also affect tool wear; a study by (Attanasio et al. 2005) indicated that applying the mist to the flank face of the tool resulted in less wear than applying mist to the rake face.

The second MQL delivery method is often referred to as single-channel. Similar to external delivery, the MQL and gas is mixed in a chamber outside the spindle, but in the case of single-channel, the MQL is delivered through the spindle to the tool holder and coolant holes in the tool and thus to the tool tip (Aoyama 2002). This method is more effective than external application in machining operations such as deep hole drilling, tapping, and reaming, where it is more difficult for coolant to reach the tool tip. This method has been found to be sensitive to the spindle rotational speed, with mist volume at the tool tip decreasing with increasing spindle speed (Dasch and Kurgin 2010).

The third major MQL delivery method is commonly referred to as dual-channel. Dual-channel is a through-spindle delivery method in which the air and oil are delivered in separate channels through the spindle and mixed at a point just before the tool holder. In a typical dual-channel system, the air flows around the lance, which has brazed heli-coils around the outside used as spacers and to make the air turbulent (Stoll et al. 2007). The MQL oil is pumped from a reservoir tank through a quick valve and rotary transmission, into a lance centered in the spindle bore (Stoll et al. 2007). The air and oil mix in a pipe nozzle located in the back of the tool holder (Stoll et al. 2007). The system delivers a fixed quantity of oil per impulse controlled by the opening time and pulse frequency of the quick valve (Stoll et al. 2007). The advantage of a dual-channel system is that it is possible to deliver more mist to the cutting zone since the oil and air do not separate due to centrifugal force in the rotating spindle (Attanasio et al. 2006). A disadvantage to dual-channel is



Minimum Quantity Lubrication (MQL) During Machining of Automotive Components, Fig. 1 Through-spindle single-channel MQL delivery system



Minimum Quantity Lubrication (MQL) During Machining of Automotive Components, Fig. 2 Through-spindle dual-channel MQL delivery system

that it is more difficult to retrofit standard CNC machines, and unlike a single-channel MQL system, it is not practical to convert back from dual-channel MQL to wet (Figs. 1 and 2).

Mist Formation and Characteristics

MQL mist is created when a small volume of oil, typically in the range of 10–300 mL/h, is mixed with air pressurized at 4–6 MPa. This combination causes a mist of oil droplets to form. The diameter of these droplets has been measured

to be in the range of 1–3 μ (Dasch and Kurgin 2010). Experiments have shown that oil volume and air pressure have an influence on droplet size, with an increase in either increasing droplet diameter (Dasch and Kurgin 2010). The authors found that the droplet size was also influenced by the delivery method, with single-channel MQL producing smaller droplets than dual-channel MQL. The average droplet diameter for single-channel was 1.7 μ , and 3.4 μ for dual-channel. Both of these measurements were significantly smaller than for wet machining, which was

tested as a comparison; the average droplet diameter for wet fluid application was $6.9\ \mu$. The droplet diameter distribution of the mist is most likely related to the volume of oil or fluid; in the case of single-channel MQL, it may be that larger droplets are produced but never reach the tip of the tool due to centrifugal force in the rotating spindle (Dasch and Kurgin 2010). Droplet diameter becomes significant when choosing mist collectors for machines in manufacturing facilities. Droplet diameter may also play a role in the mist's ability to cool and lubricate the cutting zone, with larger droplets being more desirable.

MQL Cutting Temperatures

In general, research has shown that MQL cutting temperatures are higher than for traditional wet machining (Wu and Chien 2007). Cutting temperature plays a key role in important machining metrics such as tool wear and achievable part quality. The difference may be due to the fact that wet machining fluid has been shown to have an order of magnitude higher heat transfer capacity than minimum quantity lubrication (Kurgin 2009). One reason for the superior cooling ability of wet machining coolant is that it is water-based; water is well-known to transfer heat very effectively. Also, the volume of fluid in wet machining is significantly higher. In contrast, some researchers have found that for turning steel, MQL yielded the lowest temperature compared with wet machining and dry. The authors speculate that this is due to the MQL mist's superior ability to penetrate the cutting zone (Dhar et al. 2005).

MQL Cutting Fluid Properties

As with wet machining, there are several possible liquids that can be successfully used with MQL. As stated in the introduction, the most commonly used liquids for MQL are vegetable oil and synthetic oils. There has been a fair amount of research as to the effect of oil type and properties on the effectiveness of MQL.

Suda et al. (2002) identified some important characteristics for an MQL fluid:

1. Biodegradability
2. Oxidation stability against thin film oxidation
3. Storage stability

Studies have shown that synthetic esters give better performance than vegetable oils, based on tapping tests (Suda et al. 2002). A study of the absorption capability of methyl propionate, a model ester, indicated that the oxygen present in MQL mist may promote the formation of a beneficial ester film on the tool during machining (Wakabayashi et al. 2002).

MQL Oil Volume and Air Pressure

Dual-channel MQL systems allow the user to control the rate of oil delivery and the air pressure. Oil volumes in the range of 1–500 mL/h and air pressures in the range of 4–10 bar (0.4–1.0 MPa) are typically recommended for a dual-channel system.

Many scientific studies have been done to analyze the effect of changing the volume of oil input. The importance of oil volume varies by cutting operation, machining parameters, and workpiece material. For most operations, the maximum required oil volume is 100 mL/h, with some operations successful with less than 5 mL/h. Scientific research generally indicates that “less is more” when it comes to oil volume in MQL. There has been no consistent research to indicate that more oil results in less tool wear or better surface finish; for each operation, there appears to be a minimum volume of oil required for good functionality, after which adding more oil does not improve the process (Wu and Chien 2007).

The effect of air pressure on MQL performance is not clear. In practice, the air pressure is left set at the manufacturer's recommended setting, usually between 4 and 6 MPa. The maximum air pressure setting is limited by the amount of pressurized air available in the manufacturing facility.

There is some indication that increasing air pressure reduces cutting temperature (Obikawa et al. 2006). This may be related to the fact that increasing air pressure increases the velocity of the mist, which in turn would increase its convective heat transfer coefficient, improving the cooling ability of the mist (Kurgin 2009).

MQL and Workpiece Material

Different workpiece materials have different degrees of suitability for near dry and dry machining. MQL and dry machining studies have been done for a wide variety of workpiece materials, including cast iron, steel, and aluminum. Materials that are suitable for dry cutting, such as gray iron, are readily compatible with MQL. The graphite flake inclusions found in gray iron act as a solid lubricant, partially compensating for the reduced coolant level of MQL (Klocke and Eisenblatter 1997). Most steels commonly used in engineering applications can be machined successfully with MQL (Stephenson and Filipovic). Aluminum, while it has a better machinability rating than cast iron or steel, provides a challenge for MQL machining because of its high coefficient of thermal expansion and tendency for chips to adhere to the tool (Klocke and Eisenblatter 1997). Titanium and superalloys have low

machinability ratings and are likewise difficult to machine with MQL, because of the high cutting temperatures and tendency to form built-up edge (Klocke and Eisenblatter 1997).

Alternative Environmentally Friendly Metal Cutting Fluid

Minimum quantity lubrication is one alternative to wet machining. However, it is not the only cooling technology that is the subject of current and past research.

Cryogenic cooling was studied as early as the 1960s. In this method, liquid nitrogen jet is applied to the cutting zone through a nozzle. A study compared cryogenic cooling to MQL, cold air, and dry cooling conditions when milling a titanium alloy (An et al. 2006). The researchers found that cutting zone temperatures were lower with cryogenic cooling, thus tool wear was decreased.

A recent cooling technology is known as advanced minimum quantity cooling lubrication (AMQCL). The coolant consists of solid carbon dioxide particles with a small amount of lubricant additive. The carbon dioxide particles remove heat from the cutting zone by vaporizing. The carbon dioxide also acts as a solvent, reducing the surface tension in the spray, which allows better penetration into the cutting zone. (Clarens et al. 2006) studied this topic with supercritical carbon dioxide and soybean oil. The authors found that soybean oil dissolved in supercritical CO₂ had the longest tool life compared with soybean oil microemulsion, straight soybean oil, and straight supercritical CO₂.

Cooling by heat pipe is another novel approach. A heat pipe is a passive device that transports energy (Jen et al. 2002). The heat pipe, which can be embedded in a drill, consists of a sealed pipe, a wick, and a working fluid with vapor. The heat generated by cutting enters the evaporator portion, where it causes vaporization of the cutting fluid. The vapor flows through the adiabatic transport section of the pipe to the condenser, where the heat is released to a heat sink. The condensed fluid is then pumped back to the evaporator through the wick. The heat pipe can cause a significant reduction in temperature and thus tool wear (Jen et al. 2002).

While all of these technologies have great potential, MQL is the most widely used at the present time. AMQCL can be seen as a natural evolution of MQL, while the heat pipe offers a potentially elegant way to enable dry machining for processes where it is not currently feasible. As long as there is concern for the environment and cost savings, alternatives to the current wet machining technology will continue to be explored.

Applications

The cost of traditional metal working fluids in the production of automotive components is significant; a study of several global automotive companies indicates that coolant costs within powertrain operations are in the range of 10–17% of total manufacturing costs (Stoll et al. 2007). MQL represents a cost savings due to reduced initial capital investment as compared with a centralized coolant system. Since MQL oil is used only once, ongoing costs of fluid maintenance are avoided. Potential for strict government regulations regarding coolant use and disposal is another motivation to minimize wet machining fluids. Traditional coolant may cause hazards to people in the workplace in the form of mist, which has been linked to dermatitis and respiratory issues, as well as slip and trip hazards from leaks. Because MQL uses only a small amount of vegetable or synthetic oil, these two issues are reduced or eliminated.

MQL has been successfully implemented in large-scale production of aluminum transmission components in a flexible manufacturing system (Stoll et al. 2007). MQL has also been applied to deep hole drilling of crankshaft oil holes for both steel and nodular iron crankshafts. It is often possible to create equivalent quality holes at a much faster rate with MQL than with wet machining (Stephenson and Filipovic 2006).

Key Manufacturing Considerations

MQL does not fully replicate the beneficial effects of wet machining, including chip clearing and temperature control. Methods have been developed to compensate and are described below. In addition, because the proper function of MQL depends on the precise delivery of a small amount of mist, various factors that are insignificant for wet machining become important. The following is a brief summary of key manufacturing considerations from (Stoll et al. 2007).

Cutting Tool Design

Some key design principles for cutting tools are:

1. Ensure that lubrication holes are positioned to provide oil directly to the cutting zone.
2. The outlet area of the lube holes must be less than the incoming pipe nozzle to ensure proper flow.
3. The shank of the tool should directly contact the mist nozzle; gaps result in loss of mist to the cutting tool.

Machine Design

MQL lacks the high volume of coolant that facilitates removing chips from the cutting zone, workpiece and

fixture. Allowing chips to build up in the cutting zone might reduce tool life; allowing chips to accumulate on the fixture may cause problems with part clamping and thus may be detrimental to finished part quality. Therefore, the design of a machine tool used with MQL should have enhanced ability to remove chips. For example, the use of machine beds that have steep slopes to facilitate chip removal by gravity (Filipovic and Stephenson 2006). A vacuum hood may be installed over the spindle to remove chips from the tool and part; where applicable, vertical spindles rather than horizontal spindles will encourage chips to fall away from the tool. As with machines that use a high volume of cutting fluid, the chips need to be periodically cleaned from the fixture. The use of fans to blow away chips to the machine bed, and/or spindle mounted brushes can automate the chip-clearing process (Filipovic and Stephenson 2006).

Thermal Influences

For machining workpiece materials with a high coefficient of thermal expansion, such as silicon-aluminum alloys used in the manufacture of powertrain components, it may be necessary to compensate for elevated temperatures during MQL machining. This can be accomplished through the use of temperature sensors and compensation algorithms that adjust the tool axis location and feature size based on the temperature of the part (Stoll et al. 2007). Another strategy is to optimize the sequence of machining features (in the case of a CNC operation that machines multiple features), with features of tighter tolerance machined first before the part heats up (Stoll et al. 2007).

Cross-References

- [Cutting Fluids and Their Environmental Impact](#)
- [Function of Cutting Fluids and Lubricants](#)

References

- Q.L. An, Y.C. Fu, J.H. Xu, The application of cryogenic pneumatic mist jet impinging in high-speed milling of Ti-6Al-4 V. *Key Eng. Mat.* **315–316**, 244–248 (2006)
- T. Aoyama, Development of a mixture supply system for machining with minimal quantity lubrication. *Ann. CIRP* **51**(1), 289–292 (2002)
- A. Attanasio et al., Minimal quantity lubrication in turning: effect on tool wear. *Wear* **260**, 333–338 (2006)
- A.F. Clarens, K.F. Hayes, S.J. Skerlos, Feasibility of metalworking fluids delivered in supercritical carbon dioxide. *J. Manuf. Proc.* **8**(1), 47–53 (2006)
- J.M. Dasch, S.K. Kurgin, A characterization of mist generated from minimum quantity lubrication (MQL) compared to wet machining. *Int. J. Mach. Mach. Mat.* **7**(½), 882–895 (2010)
- N.R. Dhar et al., The influence of minimum quantity lubrication (MQL) on cutting temperature, chip and dimensional accuracy in turning AISI-1040 steel. *J. Mater. Process. Technol.* **171**(1), 93–99 (2005)
- A. Filipovic, D. Stephenson, Minimum quantity lubrication (MQL) applications in automotive powertrain machining. *Mach. Sci. Technol.* **10**(1), 3–22 (2006)
- T.C. Jen et al., Investigation of heat pipe cooling in drilling applications part I: preliminary numerical analysis and verification. *Int. J. Mach. Tool Manuf.* **42**, 643–652 (2002)
- F. Klocke, G. Eisenblatter, Dry cutting. *Ann. CIRP* **46**(2), 519–526 (1997)
- S.K. Kurgin, Precision machining with minimum quantity lubrication (2009). Retrieved from UMI® 2009, <http://www.proquest.com/en-US/products/dissertations/>
- T. Obikawa, Y. Kamata, J. Shinozuka, High-speed grooving with applying MQL. *Int. J. Mach. Tool Manuf.* **46**, 1854–1861 (2006)
- A. Stoll, S. Silversson, R. Furness, Environmentally friendly and low cost manufacturing – implementation of mql machining (minimum quantity lubrication). *SAE 2007-01-1338* 2007 World Congress Detroit, MI 16–19 Apr 2007
- S. Suda et al., A synthetic ester as an optimal cutting fluid for minimal quantity lubrication machining. *CIRP Ann. Manuf. Technol.* **51.1**, 95–98 (2002)
- T. Wakabayashi et al., Tribological action and cutting performance of MQL media in machining of aluminum. *CIRP Ann. Manuf. Technol.* **56.1**, 97–100 (2007)
- C.H. Wu, C.H. Chien, Influence of lubrication type and process conditions on milling performance. *Proc. Instit. Mech. Eng. Part B J. Eng. Manuf.* **221**, 835–843 (2007)

Mixed Boundary Value Problem

- [Rigid Punch Problem with a Crack](#)

Mixed EHL

DONG ZHU

State Key Laboratory of Mechanical Transmission,
Chongqing University, Chongqing, People's Republic
of China

Synonyms

[EHL with rough surface asperity contact](#); [Mixed elastohydrodynamic lubrication](#); [Partial elastohydrodynamic lubrication](#)

Definition

Mixed elastohydrodynamic lubrication (EHL) is a mode of lubrication usually in non-conformal contacts. In mixed EHL, both hydrodynamic lubricant films and rough surface asperity contacts coexist and neither can be ignored. Since the majority of non-conformal contact components operate in a mixed EHL regime, a deep understanding of mixed EHL characteristics is

vital to performance and efficiency improvement, design optimization, and failure prevention.

Scientific Fundamentals

Introduction

EHL (elastohydrodynamic lubrication) is a mode of fluid-film lubrication in which hydrodynamic action is significantly enhanced by surface elastic deformation and lubricant viscosity increase due to high pressure. Many mechanical components operate in the EHL regime. These include various gears, rolling element bearings, cam/follower systems, vane pumps, ball screws, some metal-rolling tools, traction drives, and continuously variable transmissions. Over the last 30–60 years, fundamental EHL theories have been established, numerical solution methods developed, and various film thickness formulae derived through curve-fitting based on obtained numerical solutions. Refer to “► [Elastohydrodynamic Lubrication \(EHL\)](#),” “► [EHL Film Thickness Behavior](#),” “► [Film Thickness Formulas: Line Contacts](#),” and “► [Film Thickness Formulas: Point Contacts](#)” for details.

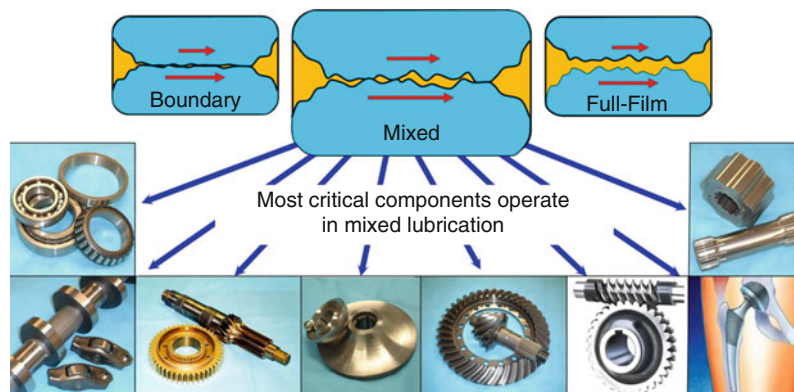
However, conventional EHL theories and various film thickness formulae were developed based on an assumption that both contacting surfaces are ideally smooth. In reality, roughness of engineering surfaces is usually of the same order of magnitude as, or greater than, the average EHL film thickness in the contact zone, so that a complete separation of the two contacting surfaces by a lubricant film is seldom achieved. In most engineering applications, hydrodynamic lubricant films and rough surface asperity contacts coexist and neither can be ignored. Figure 1 shows sketches of three major lubrication modes:

1. Full-film lubrication, in which the surfaces are separated by the lubricant film and the surface asperity contact is none or negligible
2. Mixed lubrication, where the surfaces are partially separated and asperity contacts and lubricant films coexist
3. Boundary lubrication, in which hydrodynamic action is insignificant and the load is mainly supported by the surface asperity contacts

Although mixed lubrication widely exists in many functional components, it has not been well understood in the research until recently. Challenges were mainly associated with the following two aspects:

First, conventional experimental techniques on disk machines, such as capacitance, electric resistance, and x-ray methods, are no longer satisfactory due to existence of rough surface asperity contacts. Optical interferometry, on the other hand, generally requires a transparent glass disk and a super-finished steel ball, so it is difficult to simulate asperity contacts with machined roughness, as it would require much higher resolution for film thickness measurement.

Secondly, full numerical solution of mixed EHL has been a great challenge not only due to strong nonlinearity caused by significant elastic deformation and exponentially increased viscosity at high pressure, especially under thin-film or asperity contact conditions, but also due to difficulties that one must simulate hydrodynamics and asperity contacts simultaneously in the same system. Traditionally, hydrodynamic lubrication and contact are analyzed with different model approaches. In a mixed EHL system, both hydrodynamic lubrication and contact



Mixed EHL, Fig. 1 Mixed EHL widely exists

coexist and the boundaries between two different areas are often irregular and time-dependent due to the moving rough surfaces.

Because mixed EHL is usually associated with surface roughness, great efforts have been made by researchers and engineers to develop rough surface EHL models since the 1970s. Basically, there have been two types of analysis taking into account the surface roughness effect: stochastic models and deterministic models. Early studies were focused on the stochastic models, which use a small number of statistical parameters to describe rough surface characteristics and analyze the roughness effect on contact and lubrication performance. Among the stochastic models published, the one developed by Patir and Cheng (1978) has enjoyed wide recognition. It employed an average Reynolds equation to analyze the hydrodynamics and a simplified stochastic contact model to handle the possible asperity contacts. It can be used to predict average hydrodynamic pressure and asperity contact pressure, as well as average film thickness, in the EHL conjunction, providing a simplified mathematical tool for analysis and evaluation of the overall lubrication characteristics. See “► [Stochastic Models for Rough Surface EHL](#)” for details.

It is important to note that the stochastic models deal only with the global effect of surface roughness and topography, and predict average parameter values of the rough surface EHL. In stochastic analyses, detailed information about parameter distributions and localized maximum/minimum values, which may often be critical for the study of lubrication breakdown and surface failures, are missing. Also, there have been questions and disagreements as to which stochastic model is more appropriate and which parameter(s) should be used in specific research projects or engineering applications. Since there are a variety of engineering surfaces under different operating conditions, which result in very different and complicated contact and lubrication phenomena, generally it is difficult, or even impossible, to satisfactorily describe the effects of various surface features on contact and lubrication characteristics with any simple mathematical expression using only a few stochastic parameters.

In order to obtain detailed contact and lubrication characteristics with parameter distributions, in the past 15–20 years more attention has been given to the second type of analysis – deterministic models. This type of analysis uses simplified or digitized real surface profiles in computer simulations so that statistic parameters are no longer needed in solution process. The real engineering surface topography, however, is usually very complicated,

and the problem may become strongly time-dependent due to moving surface asperities. Therefore, satisfactory deterministic solutions are much more difficult to obtain, requiring significantly more computational power and better surface measurement/analysis capability. Refer to “► [Deterministic Models of Rough Surface EHL](#)” for a brief review of model development history. This article will focus only on describing the deterministic mixed EHL modeling.

Basic Equations

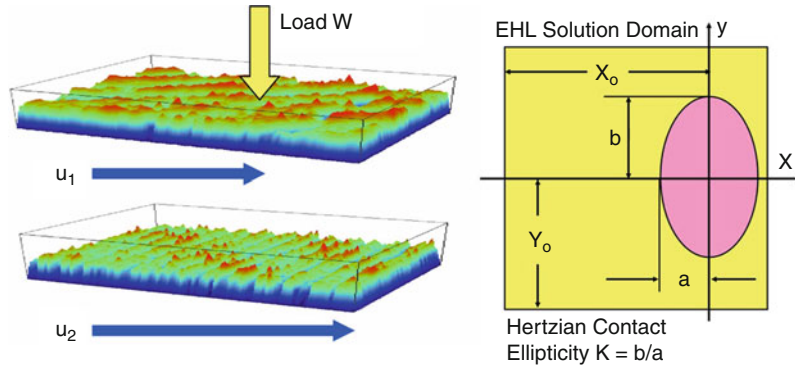
The formulation describing the EHL problems mathematically can be found in details in “► [EHL Governing Equations](#).” The following is a brief summary specifically for three-dimensional point and line contact deterministic analyses. Figure 2 gives a sketch of an EHL solution domain for deterministic simulation, in which two arbitrary rough surfaces are running through the EHL conjunction at any rolling and relative sliding velocities. The pressure within the entire domain is governed by the Reynolds equation below:

$$\begin{aligned} \frac{\partial}{\partial x} \left(\frac{\rho h^3}{12\eta} \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{\rho h^3}{12\eta} \frac{\partial p}{\partial y} \right) \\ = \frac{u_1 + u_2}{2} \frac{\partial(\rho h)}{\partial x} + \frac{\partial(\rho h)}{\partial t} \end{aligned} \quad (1)$$

In which the lubricant can be Newtonian or non-Newtonian, and all the parameters are defined in the nomenclature. The x-coordinate is chosen to coincide with the rolling direction. A boundary condition of pressure $p = 0$ at the edges of the solution domain as well as a normal cavitation condition must be satisfied when solving this equation. For a point contact EHL problem with the two moving rough contacting surfaces, the local lubricant film thickness/gap is time-dependent and can be expressed as

$$\begin{aligned} h(x, y, t) = h_o(t) + \frac{x^2}{2R_x} + \frac{y^2}{2R_y} + V(x, y, t) \\ + \delta_1(x, y, t) + \delta_2(x, y, t) \end{aligned} \quad (2)$$

Note that the first term on the right-hand side represents a normal approach between the two bodies, and the second and third terms original macro contact geometry. $V(x, y, t)$ is the surface elastic deformation, and $\delta_1(x, y, t)$ and $\delta_2(x, y, t)$ are three-dimensional original roughness profiles for surfaces No. 1 and No. 2, respectively. Because both surfaces are moving along the x-direction, δ_1 and δ_2 , are time-dependent, and can be given by



Mixed EHL, Fig. 2 Sketch of solution domain for point contact EHL

$$\begin{aligned}\delta_1(x, y, t) &= s_1(x - u_1 t, y) \\ \delta_2(x, y, t) &= s_2(x - u_2 t, y)\end{aligned}\quad (3)$$

where $s_1(x, y)$ and $s_2(x, y)$ are surface profiles, typically in the form of measured discretized roughness height data matrices. For line contact problems, the macro contact geometry is two-dimensional while the surface roughness may still be three-dimensional, so that (2) can be written as

$$h(x, y, t) = h_o(t) + \frac{x^2}{2R_x} + V(x, y, t) + \delta_1(x, y, t) + \delta_2(x, y, t) \quad (4)$$

The surface elastic deformation is computed through the following integral:

$$V(x, y, t) = \frac{2}{\pi E'} \iint_{\Omega} \frac{p(\xi, \zeta)}{\sqrt{(x - \xi)^2 + (y - \zeta)^2}} d\xi d\zeta \quad (5)$$

The lubricant viscosity is assumed to be dependent on pressure, and one of the commonly used viscosity equations is the Barus law given below:

$$\eta = \eta_o \exp(\alpha p) \quad (6)$$

The density of lubricant is also a function of pressure, commonly calculated by:

$$\rho = \rho_o \left(1 + \frac{0.6 * 10^{-9} p}{1 + 1.7 * 10^{-9} p} \right) \quad (7)$$

The applied load is balanced by the integral of the pressure over the entire solution domain Ω , i.e.,

$$w(t) = \iint_{\Omega} p(x, y, t) dx dy \quad (8)$$

All the equations given above form a strong nonlinear system, which can be discretized and then solved with an

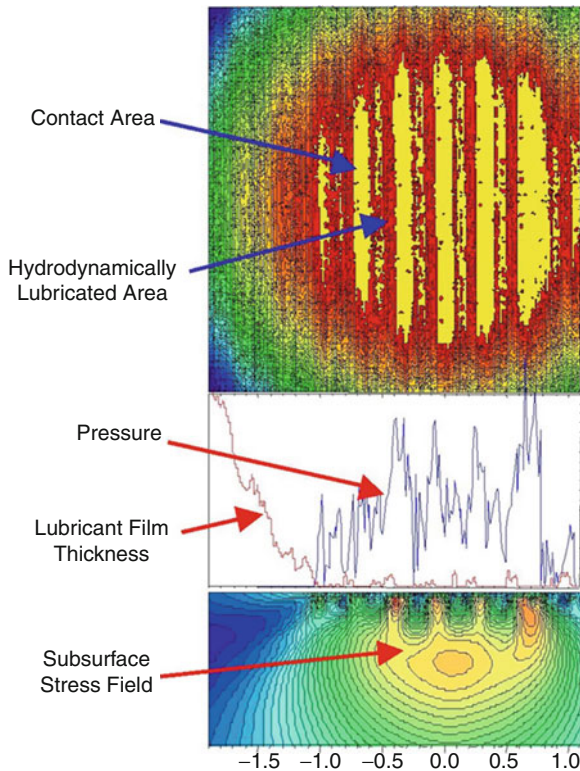
iterative procedure. The numerical solution techniques are described in “►EHL, Full Numerical Solution Methods” so they will not be repeated here.

Two Solution Approaches for Mixed EHL

In mixed lubrication there are two different areas, as illustrated in Fig. 3. In the hydrodynamically lubricated areas the contacting surfaces are separated by lubricant films, while in the asperity contact areas there is no significant lubricant film and the surfaces are in direct contact, but boundary films consisting of a few layers of molecules may exist. When simulating the mixed EHL, the key is how to model these two different types of areas simultaneously.

Basically, there have been two approaches: the first is to use a unified equation system and solution method for both the lubricated areas and asperity contacts simultaneously, and the second is to use separate models for lubrication and contact. Jiang et al. (1999), presented the first separate approach for point contact mixed EHL with machined roughness. In the same year the first unified approach for point contacts with machined three-dimensional roughness was published by Zhu and Hu (1999) then by Hu and Zhu (2000). Other published unified solutions include those by Holmes et al. (2005), and Li and Kahraman (2009). The separate approach was also employed by Zhao et al. (2001), Popovici et al. (2004), and others, mainly for start-up and slow-down problems, in which the borders between the lubricated and contact areas are relatively simple so the boundary conditions are easy to handle.

Traditionally, hydrodynamic lubrication has been modeled with the Reynolds equation, while the surface contact analyzed by using solid mechanics without considering hydrodynamics. However, linking two different models and properly handling boundary conditions between the two types of areas could be difficult, as the



Mixed EHL, Fig. 3 Hydrodynamically lubricated and contact areas

asperity contact size and the border shape are random, irregular, and strongly time-dependent due to moving machined roughness. It is highly preferred to employ a unified model for simulating both the hydrodynamic lubrication and the contact in one system, so that dealing with complicated boundary conditions can be avoided. Since dry contact is nothing but a special case of lubricated contact under extreme conditions (such as extremely low speed, extremely low viscosity, or high pressure concentration in small asperity contacts), there is no reason one cannot use lubrication equations to solve dry contact problems, if the numerical solver is sufficiently robust. A semi-system approach for the EHL solution, originally presented by Ai (1993) and modified by Zhu and Ai (1997), Hu and Zhu (2000), and others, appears to be able to handle very thin EHL films and asperity contacts, capable of solving mixed EHL problems with machined roughness by using a unified equation system and numerical solution approach. Refer to “► [Semi-System Approach](#)” and “► [EHL, Full Numerical Solution Methods](#)” for details.

General Characteristics of Mixed EHL

First, before describing the mixed EHL characteristics, it is important to note that previously defined parameters for smooth surface EHL, such as the central and minimum film thicknesses, are no longer appropriate. The minimum film thickness is always zero whenever/wherever there is a local asperity contact, and the central film thickness may fluctuate greatly from one moment to another due to moving roughness. New parameters are introduced as follows in order to describe mixed EHL behavior:

Average Film Thickness h_a : For point contacts, h_a is calculated within a certain area around the center of contact, typically within $\frac{1}{2}$ Hertzian radius of the normalized contact zone, so that a sufficiently large area can be included for calculating the average value, but possible edge effect can be avoided. For line contacts, the average value is calculated within $\frac{1}{2}$ Hertzian contact width on each side along the centerline. When the surfaces are smooth, h_a is almost the same as the central film thickness h_c defined conventionally.

Film Thickness (λ) Ratio: In the last 30+ years the term of λ ratio has been widely used, but defined in different ways, so it should be clarified. For mixed EHL it is now suggested that λ ratio is defined as the ratio of average film thickness (defined above) to the composite RMS roughness, $\lambda = h_a/\sigma$. This definition well describes the global lubrication effectiveness in rough surface contacts.

Contact Load Ratio W_c and Contact Area Ratio A_c : If the average EHL film thickness is not much greater than the composite RMS roughness, say the λ ratio is smaller than 1.0~1.2 or so, surface asperity contacts would usually take place, and a certain portion of load is supported by the asperity contacts. The contact area ratio, A_c , is defined as the ratio of total asperity contact area to the area of Hertzian contact zone. The contact load ratio, W_c , is the load supported by asperity contacts divided by the total load. Note that A_c and W_c are correlated but can be different quantitatively. For full-film lubrication with no significant asperity contact, both A_c and W_c are zero or nearly zero. As the film thickness decreases and the asperity contacts become more significant, A_c and W_c increase. For dry contact W_c will reach its maximum value of 1.0, but A_c may be less than 1.0, unless all the asperities are completely flattened.

Coefficients of Hydrodynamic Friction, f_h , and Contact Friction, f_c : Friction in mixed lubrication usually consists of two parts: the hydrodynamic friction is caused by lubricant shearing in the EHL films while the contact friction due to boundary lubrication or dry contact in asperity contact areas. The total friction coefficient, f , is

often dominated by the contact friction, as it is usually much greater than the hydrodynamic one.

In order to demonstrate the general characteristics of mixed EHL, a set of cases are analyzed and results illustrated in Fig. 4. Two machined (shaved) surfaces are running through a circular EHL contact, and the entraining speed varies in a wide range from 50 m/s down to 1×10^{-9} m/s, so that the solutions cover the entire transition from the full-film and mixed EHL down to boundary lubrication. For comparison purposes, a smooth surface solution at $U = 50$ m/s is also given in Fig. 4. It can be seen that at a high speed of 50 m/s the surfaces are fully separated. As the speed decreases, the average lubricant film thickness becomes smaller and an increasing number of asperity contacts can be observed. When the speed is low, say below 0.1 m/s, the asperity contact areas (shown as light-yellow areas in Fig. 4) become more than 50% of the total area of Hertzian contact zone. If the speed is extremely low, e.g., 1×10^{-9} m/s, the asperity contact becomes dominant and the hydrodynamic lubrication vanishes.

Figure 5 shows general trends of lubrication transition. It is observed that as the entraining speed increases, the average film thickness and λ ratio increase significantly, while the contact load ratio and friction decrease due to reduced asperity contacts. When the λ ratio is smaller than $0.02 \sim 0.03$ or so, more than 85% of load is supported by the asperity contacts, and the friction coefficient, f , is very close to that of the boundary lubrication, f_b . This can be considered as boundary lubrication (or dry contact). On the other hand, if the λ ratio is greater than $1.0 \sim 1.2$ or so, the contact load ratio becomes zero or negligible, demonstrating a full-film EHL status with a thick lubricant film and much reduced friction.

It has been noticed that the results obtained from the present deterministic model show the same basic trends as those predicted by stochastic models, but quantitative differences exist. For example, it was believed based on stochastic analyses that the mixed lubrication is in a range of λ ratio from 0.5 to 3.0. Actually, the stochastic models, such as those by Patir and Cheng (1978), and Zhu and Cheng (1988), could not analyze cases at λ ratios below 0.5 due to intrinsic limitations of the models, which did not properly consider asperity deformation and interaction. The present deterministic model is able to handle very deep asperity penetration at very small λ ratios, yielding more reasonable results that agree well with available experiments. A comparison between the model prediction and experimental observation for relationship of the contact area ratio vs. the λ ratio is given in Fig. 6. It can be seen that the predicted λ ratio range for the mixed lubrication is

well verified by experiments. Similar results have also been presented by Guangteng and Spikes (1997).

Friction in Mixed EHL

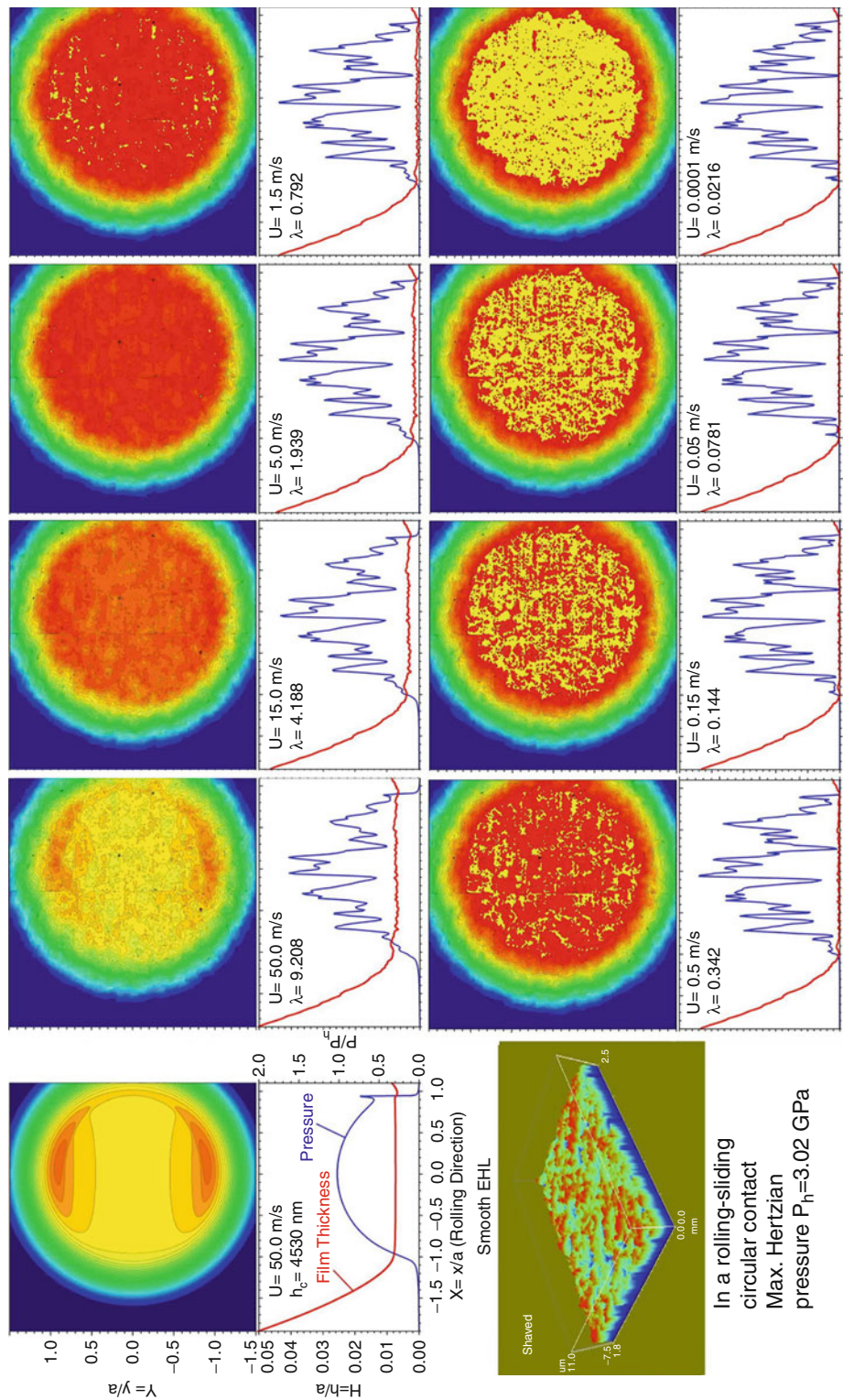
Friction in mixed EHL is of great importance in practice, as it is directly related to efficiency, energy saving, and system thermal behavior. It is understood that even a small percentage of energy loss due to friction may cause considerable temperature increase in the system, and this may in turn lead to a significant reduction of lubricant viscosity, which may negatively affect the lubrication performance, resulting in further increase in friction due to reduced EHL film thickness and increased asperity contact severity. This unfavorable chain reaction may cause quick lubrication breakdown and early failure of the components.

Friction in mixed EHL consists of two parts. In hydrodynamically lubricated areas the friction is generated from lubricant shearing in the EHL films. Since the pressure and the shear rate are typically very high in the Hertzian contact zone, the Newtonian fluid model is no longer valid, and an appropriate non-Newtonian fluid model is needed. A common practice is to use the viscous-elastic model proposed by Bair and Winer (1978) given as follows:

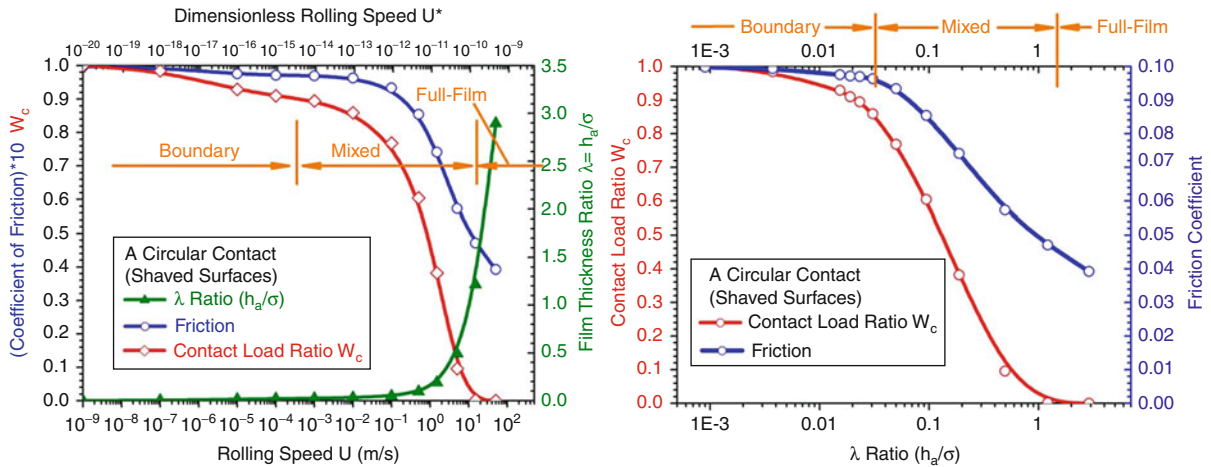
$$\dot{\gamma} = \frac{\dot{\tau}}{G_x} - \frac{\tau_L}{\eta} \ln \left(1 - \frac{\tau}{\tau_L} \right) \quad (9)$$

where τ_L and G_x are the limiting shear stress and limiting elastic modulus, respectively, which are properties of the lubricant and also functions of pressure and temperature. Generally, advanced synthetic lubricants exhibit low friction behavior because of their low limiting shear stress, while traction fluids for continuous variable transmissions are specifically formulated to have high limiting shear stress. By solving the above nonlinear equation based on given τ_L and G_x , one can calculate the hydrodynamic friction through integration of the shear stress in lubricated areas.

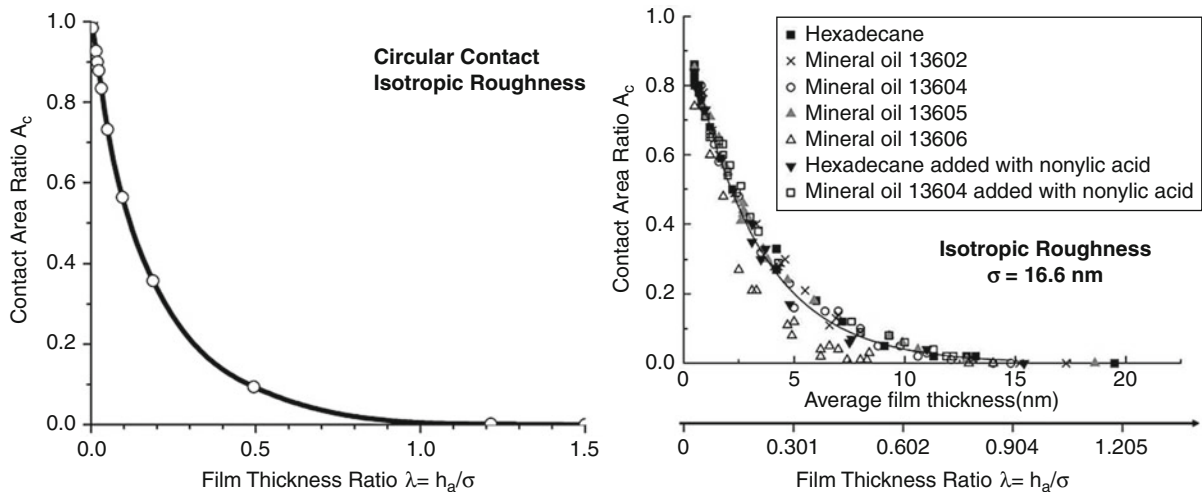
In asperity contact areas the friction is usually much higher, determined by boundary lubrication properties. Boundary friction is to date difficult to model as it involves complicated interfacial physics and chemistry, but it is usually easy to estimate experimentally. Boundary friction coefficient f_b is often in a range of 0.07–0.15 for lubrication systems with commonly used steels and industrial lubricants. If f_b is provided as an input, the total friction in the mixed EHL can be predicted, as described in detail by Martini et al. (2007). It is obvious that the total friction is greatly affected by how much asperity contact



Mixed EHL, Fig. 4 Transition from full-film and mixed EHL to boundary lubrication (Circular contact, $G^* = 5817.8$, $W^* = 4.9083 \times 10^{-5}$, $U^* = 1.6197 \times 10^{-15}$, $U^* \sim 8.0987 \times 10^{-10}$)



Mixed EHL, Fig. 5 General trends for cases shown in Fig. 4 (f_b is assumed to be 0.1)



Mixed EHL, Fig. 6 Comparison of predicted contact area ratios (Left) with experimental data from Luo and Liu, 2006 (Right)

exists in the mixed EHL system. Generally, the larger the contact load ratio W_c , the greater the friction.

The friction curves given in Fig. 5 are obtained through the above-mentioned approach, demonstrating frictional behavior over the entire transition from the full-film and mixed EHL to boundary lubrication. This type of friction curve is commonly called the “Stribeck curve” in the literature. Generally, in the boundary lubrication friction is high and quite constant, as the hydrodynamic action is negligible. As the speed or λ ratio increases, it enters the mixed EHL regime and the friction quickly decreases due to enhanced hydrodynamics

and reduced asperity contacts. In the full-film EHL region the friction is usually low, dependent mainly upon the lubricant rheological properties.

It is understood based on the mechanism of mixed EHL friction described above that the friction can be significantly reduced by the following three major means, which are the technologies currently available:

1. Use better finished surfaces with considerably lower RMS roughness, and/or optimize design and operating conditions, in order to increase the λ ratio and minimize the surface asperity contacts.

2. Reduce boundary friction coefficient by employing advanced boundary additives and low-friction materials and coatings. This is very effective especially for heavy-duty components in which asperity contacts may be severe.
3. Choose advanced lubes, such as synthetic oils, with low limiting shear stress, so that the hydrodynamic component of friction due to lubricant shearing can be reduced.

Figure 7 shows some analyzed examples demonstrating the effectiveness of the friction reduction means stated above. It can be seen that by using these three techniques combined, the total friction can be reduced significantly, say by up to 70%, from the situation of shaved surface finish with regular mineral oil and no advanced boundary additive (that widely exists currently in practice) to improved condition with honed or polished surfaces and low-shear synthetic lubes plus advanced additives with reduced f_b .

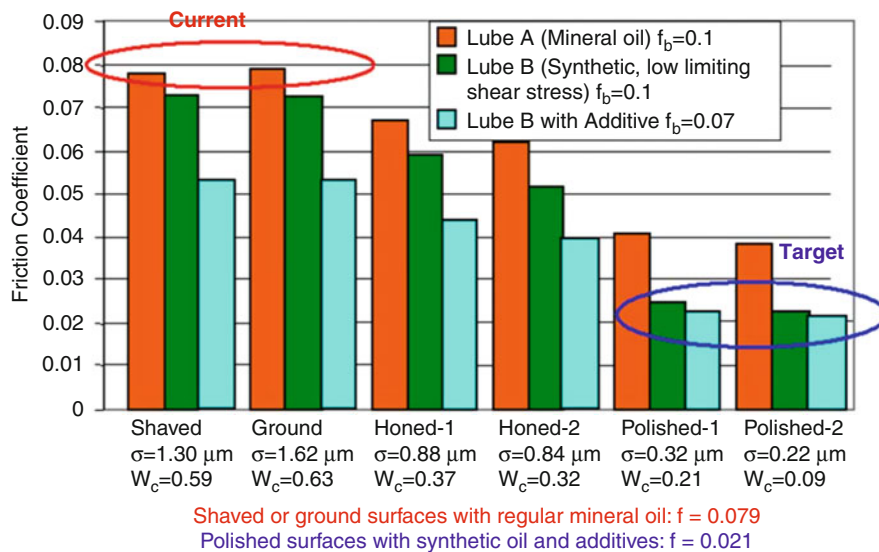
Effect of Surface Roughness and Topography

Mixed EHL characteristics are usually influenced significantly by surface roughness and topography, so this is a topic of great importance in both lubrication theory and engineering practice. The roughness effect mainly involves three aspects: (1) roughness height, (2) rough surface topography/textures, and (3) rough surface orientation.

Key Applications

Mixed EHL is a mode of lubrication in which most functional components with non-conformal contacts operate. These components include various gears, rolling bearings, cams and followers, hydraulic vane pumps, ball screws, traction drives, continuous variable transmissions, and metal rolling tools. These types of components usually transmit substantial power and motion, and they often carry heavy loads. Due to high stress concentrated in small contact areas and severe asperity contacts at the interface, these components are often vulnerable points in machinery. Inadequate lubrication may cause early surface failures such as excessive wear, scuffing, and pitting due to contact fatigue. A good understanding of the mixed EHL is vital to products' design optimization, performance improvement, efficiency and durability maximization, and failure prevention and to environmental protection and energy conservation.

There are two main types of models for analyzing the surface roughness effect in mixed EHL: stochastic and deterministic. Stochastic approaches are relatively simple and efficient, but they only provide mean values of lubricant film thickness and hydrodynamic and contact pressures, demonstrating global trends of lubrication performance and transition. Deterministic models are able to provide more detailed information about the film thickness and pressure fluctuations. As computer technology has advanced, deterministic models have



Mixed EHL, Fig. 7 Effects of surface finish, lubricant and boundary additive on friction ($U = 7.5$ m/s, $S = 0.25$, Circular Contact, $P_h = 2.155$ GPa)

continuously been improved. Recently developed deterministic mixed EHL models are capable of handling severe operating conditions and simulating the entire transition from full-film and mixed EHL to boundary lubrication or even dry contact, with fully digitized, three-dimensional machined rough surfaces as input data. It has been proven that the unified mixed EHL model can provide detailed transient parameter distributions, which are vital for the investigation of lubrication effectiveness, lubrication breakdown, and surface failures. This type of model can be applied as a useful tool in research projects and engineering design analyses.

Nomenclature

a	Semi-axis of Hertzian contact ellipse in rolling direction, or radius of Hertzian contact circle, or half-width of Hertzian contact for a line contact
A_c	Contact area ratio
b	Semi-axis of Hertzian contact ellipse in the direction perpendicular to rolling
E'	Effective elastic modulus
f	Friction coefficient
f_b	Boundary friction coefficient
f_h	Hydrodynamic friction coefficient
$G^* = \alpha E'$	Dimensionless material parameter
h	Local film thickness (or gap)
h_a	Average film thickness (or average gap)
l_e	Effective length of line contact
p	Pressure
P_h	Maximum Hertzian pressure
R_q	Root mean square (RMS) surface roughness
R_x, R_y	Effective radii of curvature in x- and y-directions, respectively
$S = (u_2 - u_1)/U$	Slide-to-roll ratio
t	Time
$U^* = \eta_0 U / (E' R_x)$	Dimensionless speed parameter
$U = (u_1 + u_2)/2$	Rolling velocity (or entraining velocity)
u_1, u_2	Velocities of Surface 1 and Surface 2, respectively
V	Surface elastic deformation
W^*	Dimensionless load parameter, point contact: $w/(E' R_x^2)$; line contact: $w/(E' R_x l_e)$
w	Load
W_c	Contact load ratio (load supported by surface contact divided by total load)

x, y	Coordinates (x is chosen to be parallel to rolling direction)
α	Pressure-viscosity exponent used in $\eta = \eta_0 \exp(\alpha p)$
δ_1, δ_2	Roughness height of surfaces 1 and 2, respectively
η, η_0	Viscosity and viscosity under ambient condition, respectively
ρ, ρ_0	Density and density under ambient condition, respectively
$\lambda = h_a / \sigma$	Film thickness ratio
$\sigma = (R_{q1}^2 + R_{q2}^2)^{0.5}$	Composite RMS roughness

Cross-References

- [3D Line Contact EHL](#)
- [Deterministic Models of Rough Surface EHL](#)
- [EHL Film Thickness Behavior](#)
- [EHL, Full Numerical Solution Methods](#)
- [EHL Governing Equations](#)
- [Elastohydrodynamic Lubrication \(EHL\)](#)
- [Film Thickness Formulas: Line Contacts](#)
- [Film Thickness Formulas: Point Contacts](#)
- [Semi-System Approach](#)
- [Stochastic Models for Rough Surface EHL](#)

References

- X. Ai, Numerical analyses of elastohydrodynamically lubricated line and point contacts with rough surfaces by using semi-system and multigrid methods, Ph.D. Thesis, Northwestern University, Evanston, Illinois, USA, 1993
- S. Bair, W.O. Winer, Rheological response of lubricants in EHD contacts, in *Proceedings of the 5th Leeds-Lyon Symposium on Tribology* (The University of Leeds, England, 1978), pp.162–169
- G. Guangteng, H.A. Spikes, The control of friction by molecular fractionation of base fluid mixtures at metal surfaces. *Tribol. Trans.* **40**, 461–469 (1997)
- M.J.A. Holmes, H.P. Evans, R.W. Snidle, Analysis of mixed lubrication effects in simulated gear tooth contacts. *J. Tribol.* **127**, 61–69 (2005)
- Y.Z. Hu, D. Zhu, A full numerical solution to the mixed lubrication in point contacts. *J. Tribol.* **122**, 1–9 (2000)
- X. Jiang, D.Y. Hua, H.S. Cheng, X. Ai, S.C. Lee, A mixed elastohydrodynamic lubrication model with asperity contact. *J. Tribol.* **121**, 481–491 (1999)
- S. Li, A. Kahraman, A mixed EHL model with asymmetric integrated control volume discretization. *Tribol. Inter.* **42**, 1163–1172 (2009)
- J.B. Luo, S. Liu, The investigation of contact ratio in mixed lubrication. *Tribol. Inter.* **39**, 409–416 (2006)
- A. Martini, D. Zhu, Q. Wang, Friction reduction in mixed lubrication. *Tribol. Lett.* **28**, 171–181 (2007)
- N. Patir, H.S. Cheng, Effect of surface roughness orientation on the central film thickness in EHD contacts, in *Proceedings of the 5th Leeds-Lyon Symposium on Tribology* (The University of Leeds, England, 1978), pp.15–21

- G. Popovici, C.H. Venner, P.M. Lugt, Effects of load system dynamics on the film thickness in EHL contacts during start up. *J. Tribol.* **126**, 258–266 (2004)
- J.X. Zhao, F. Sadeghi, M.H. Hoeprich, Analysis of EHL circular contact start up: part I—mixed contact model with pressure and film thickness results. *J. Tribol.* **123**, 67–74 (2001)
- D. Zhu, X. Ai, Point contact EHL based on optically measured three-dimensional rough surfaces. *J. Tribol.* **119**, 375–384 (1997)
- D. Zhu, H.S. Cheng, Effect of surface roughness on the point contact EHL. *J. Tribol.* **110**, 32–37 (1988)
- D. Zhu, Y.Z. Hu, The study of transition from full film elastohydrodynamic to mixed and boundary lubrication, *The Advancing Frontier of Engineering Tribology, in Proceedings of the 1999 STLE/ASME H.S. Cheng Tribology Surveillance*, Park Ridge, Illinois, USA, (1999), pp.150–156

Mixed EHL in Gears

DONG ZHU

State Key Laboratory of Mechanical Transmission,
Chongqing University, Chongqing, People's Republic
of China

Synonyms

Application of mixed EHL in gears; Gears and mixed lubrication

Definition

Most gears operate in a mixed elastohydrodynamic lubrication (EHL) regime, in which both hydrodynamic lubricant films and rough surface asperity contacts coexist. A good understanding of mixed EHL characteristics is vital to gear performance and efficiency improvement, design optimization, and failure prevention.

Scientific Fundamentals

Introduction

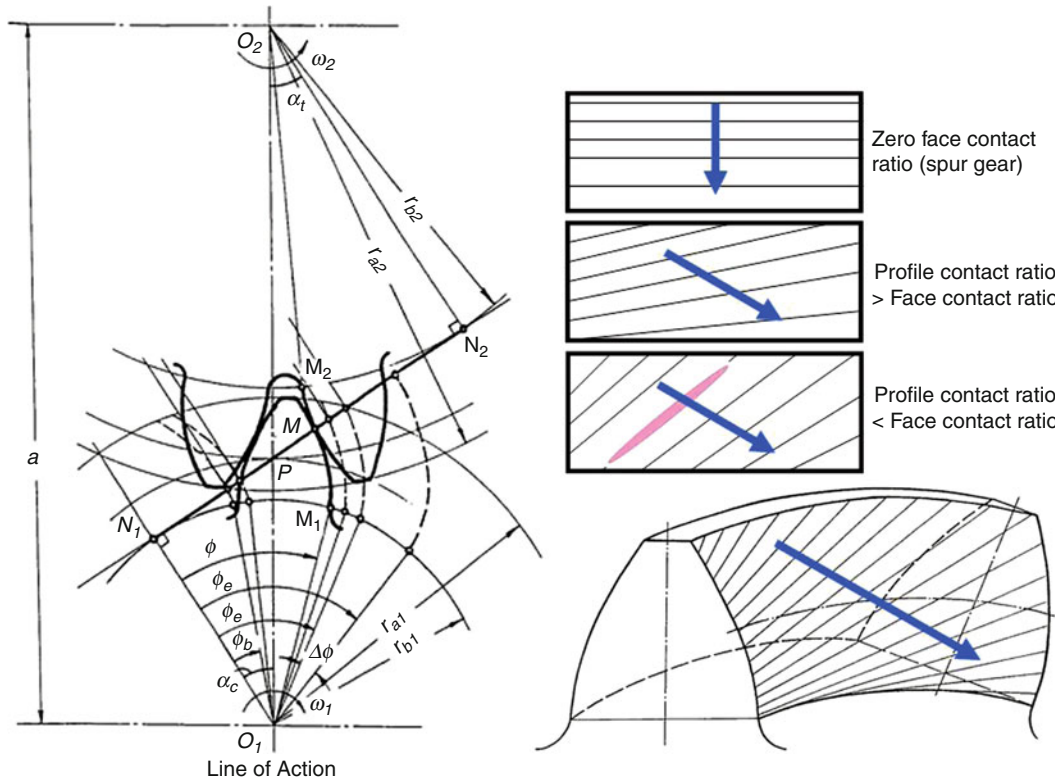
Gears are key components widely used in various vehicles and machineries for transmitting power and motion through tooth surface contacts. Although practical application of gears can be traced back to thousands of years ago, e.g., in ancient Egypt, Greece, and China, the fundamental mechanisms of gear lubrication was not understood until quite recently. The following are some basic characteristics of gear lubrication that have presented difficulties to researchers and engineers:

1. Interaction between gear teeth is carried out through non-conformal surface contacts, which are often lubricated. Non-conformal (or concentrated) contact

can be classified as line contact (often found in spur, helical, and straight-bevel gears and most worm gears) and point contact (in hypoid and spiral bevel gears, helical gears with crossed axes, and circular-arc (Novikov) gears). Localized high contact pressure in the small contact zone between the gear teeth can be as high as 1–4 GPa, causing significant surface deformation and viscosity increase. Therefore, gears are typical components operating in the EHL regime.

2. Commonly used machining processes for finishing gear tooth surfaces include hobbing, milling, shaping, rolling, shaving, broaching, grinding, lapping, honing, and polishing. Most processes produce tooth flank surfaces of roughness R_a greater than 0.5–1.0 micron, while a small percentage of gears may be better finished by precise grinding, honing, or polishing to $R_a = 0.15\text{--}0.4\text{ }\mu\text{m}$ or so. It is often difficult and costly to obtain significantly better tooth surface finish due to complicated gear geometry. The surface roughness, therefore, is usually of the same order as, or greater than, the possible EHL film thickness in tooth contacts. A complete separation of the tooth surfaces by a hydrodynamic lubricant film is seldom achieved. In most cases, the lubricant films and rough surface asperity contacts coexist, and neither can be ignored. When studying gear lubrication, one has to consider both simultaneously in the same system of mixed EHL.
3. When a pair of gears is engaged, the contact between two teeth moves in a specific way. For a spur-driven gear, for example, the location of the tooth contact line moves from tip to root on the tooth surface, while for a helical gear it moves in a diagonal angle, as shown in Fig. 1. In worm or hypoid gears the contact and its movement are much more complicated. During the movement the tooth load, the rolling and relative sliding velocities, and the radii of curvature of tooth surfaces are all changing constantly. The number of pairs of teeth that share the load keeps changing as well. Machining and assembly errors/misalignments and tooth geometric modifications may bring in more complexity especially at high speeds. The contact and lubrication conditions, therefore, are basically transient, influenced by gear dynamics and many other factors.

For the reasons mentioned above, the contact and lubrication mechanisms of gears, even of the simpler types such as spur and helical, are very complicated. In practice, today engineers still largely rely on empirical data and experience. Model-based comprehensive lubrication design and analysis tools simulating all the dynamic and



Contact line on gear tooth surface moves during engagement.

Mixed EHL in Gears, Fig. 1 Sketch of tooth contact in a pair of helical/spur gears

mixed EHL mechanisms are not yet available. Simplified analyses of various types, however, can be found in the literature. One of the remarkable early efforts was by H. M. Martin, who published his hydrodynamic lubrication analysis for line contacts in spur gears in 1916 (Martin 1916). In his study, a pair of gear teeth was simplified to two parallel smooth rigid cylinders lubricated by an incompressible isoviscous Newtonian fluid. It was found, however, that the lubricant film thickness between gear teeth estimated by Martin's model was extremely small and disagreed with engineering observations. Reasonable lubricant film thickness in gear tooth contacts was first predicted by Dowson and Higginson (1966) based on line contact EHL theory under smooth surface, steady-state, and isothermal assumptions. Since transient EHL models are complicated, Gu (1973) proposed a criterion that can justify the use of a quasi-steady solution for analyzing gear lubrication. Estimating film thickness in gears by using steady-state EHL formulae (such as that by Dowson and Higginson (1961)) at chosen critical contact locations has become a common practice since then. Predicted film thickness is found to be a good

reference in gear design. Later, Wang and Cheng (1981) developed a numerical solution of EHL film thickness, friction, and surface temperature for an entire meshing cycle in spur gears, considering dynamic and thermal effects, but surface roughness was neglected. Zhu and Cheng (1991) extended the analysis to a three-dimensional (3D) one for helical gears, but dynamic effect on load distribution between pairs of teeth was ignored for simplicity.

Rapidly developing computer and information technologies have fueled significant advancements in thin-film and mixed EHL theory and practice since the late 1990s. Deterministic solutions for mixed EHL with real machined rough surfaces using a unified numerical approach have been obtained by Zhu and Hu (1999), Hu and Zhu (2000), and Holmes et al. (2005), and others. Newly developed numerical approaches appear to be capable of simulating the entire transition from full-film and mixed EHL down to dry contact under severe operating conditions. Also, a 3D line contact mixed EHL model has been developed by Ren et al. (2009). This is of great importance to gear lubrication studies, because line

contacts widely exist in various gears but machined gear tooth surface topography is usually three-dimensional. A 3D deterministic model is needed to simulate the effects of surface roughness and topography on contact and lubrication, although the macro tooth contact geometry may be simplified to two-dimensional (2D). Based on the development of mixed EHL mentioned above, Martini et al. (2007) analyzed mixed EHL frictional behavior in gears and proposed practical means for friction reduction. Zhu (2009) developed a comprehensive analysis for mixed EHL performance and friction in spur and helical gears. The mixed EHL model is able to predict asperity contact severity between gear teeth, which can be directly correlated with surface failures, such as excessive wear (see Zhu et al. 2007) and pitting due to contact fatigue (see Zhu et al. 2009).

Once again, gear lubrication is a complicated topic. This essay will focus on the mixed EHL characteristics in spur and helical gears without in-depth discussion on gear dynamics.

Gear Geometry and Kinematics

Spur and helical gears are commonly used in industry. During a meshing cycle, the radii of curvature and the rolling (or entraining) and sliding velocities at the tooth contact location keep changing constantly. If ϕ is the rotation angle specifying the location of tooth contact on a transverse plane, as shown in Fig. 1, the entraining velocity U and sliding velocity V at that location can be given by

$$U = 0.5\omega_1 \cos \beta_b \left[r_{b1}\phi + \frac{Z_1}{Z_2} (A \sin \alpha_t - r_{b1}\phi) \right]$$

$$V = \omega_1 \cos \beta_b \left[r_{b1}\phi - \frac{Z_1}{Z_2} (A \sin \alpha_t - r_{b1}\phi) \right]$$

where ω_1 is the rotational speed of pinion, α_t the transverse pressure angle, A the center distance between the pinion and gear, β_b the base helix angle, and r_{b1} the radius of base cylinder of pinion; Z_1 and Z_2 are numbers of teeth of pinion and gear, respectively. The principal radius of curvature in the direction perpendicular to the contact line can be calculated from

$$\rho = r \sin \alpha_r / \sin \lambda_b$$

where $r = (x^2 + y^2)^{0.5}$, $\alpha_r = \cos^{-1}(r_b/r)$, $\lambda_b = 0.5\pi - \beta_b$, and x and y are coordinates of the spot on the contact line. Detailed geometric calculations can be found in the paper by Zhu and Cheng (1991). The effective radius of curvature in the rolling direction is defined as

$$R_x = \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} \right)^{-1}$$

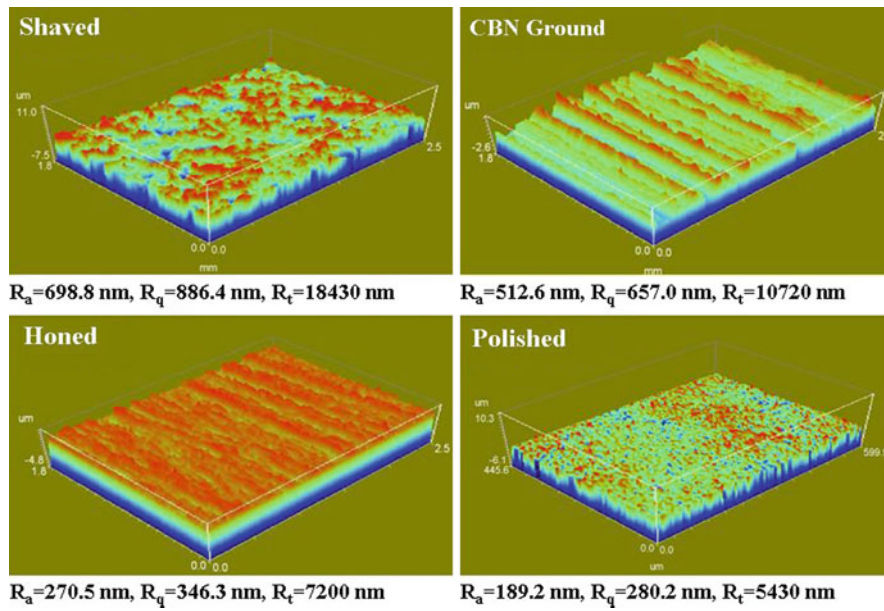
where ρ_1 and ρ_2 are the radii of curvature of pinion and gear at the contact location, respectively. Note that U , V , and R_x are key parameters for lubrication study. They vary during the entire mesh cycle and also along the contact line if helix angle is not zero.

3D Line Contact Mixed EHL Model

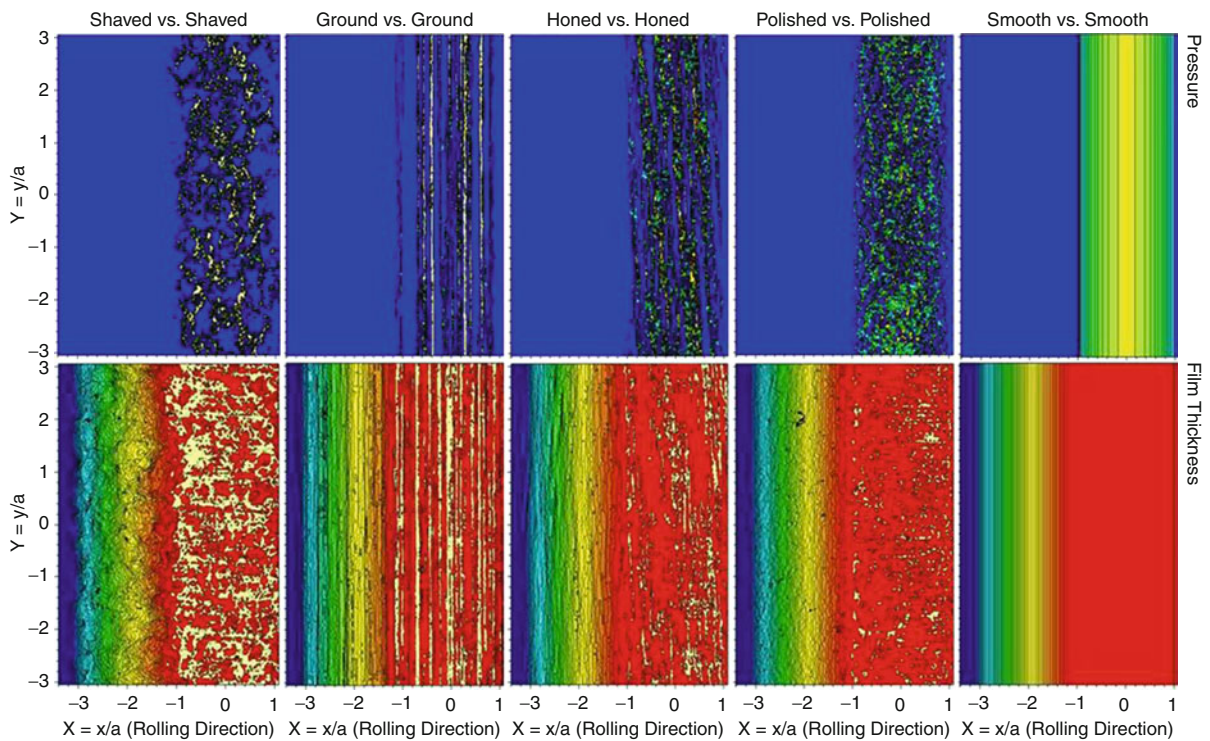
As is well known, spur and helical gears are typical line contact components, for which the contact geometry can be simplified to 2D, represented by two parallel cylinders. In reality, however, gears are machined and tooth surface topography is usually 3D. Figure 2 shows some sample 3D surface profiles optically measured from gear teeth. They are made by commonly used machining processes for gears, such as shaving, grinding, honing, and polishing. As 2D line contact EHL models cannot simulate 3D EHL characteristics caused by 3D surface topography, a 3D line contact EHL model has been developed. The model numerically solves an EHL equation system, which consists of the Reynolds equation for hydrodynamics, an elasticity equation for surface deformation, a geometric equation for contact geometry and film thickness, and lubricant rheology equations for viscosity and density. Although the contact line should be infinitely long in the y -direction (or much longer than its width), the solution domain has to be cut to a finite length based on an assumption that 3D roughness and resulting pressure and film thickness distributions are all periodically repeated along the contact length direction. When calculating the surface deformation in the solution domain, which results from the pressure distribution not only within the domain but outside the domain as well, a mixed FFT algorithm is employed. Refer to “► 3D Line Contact EHL,” or the paper by Ren et al. (2009) for details.

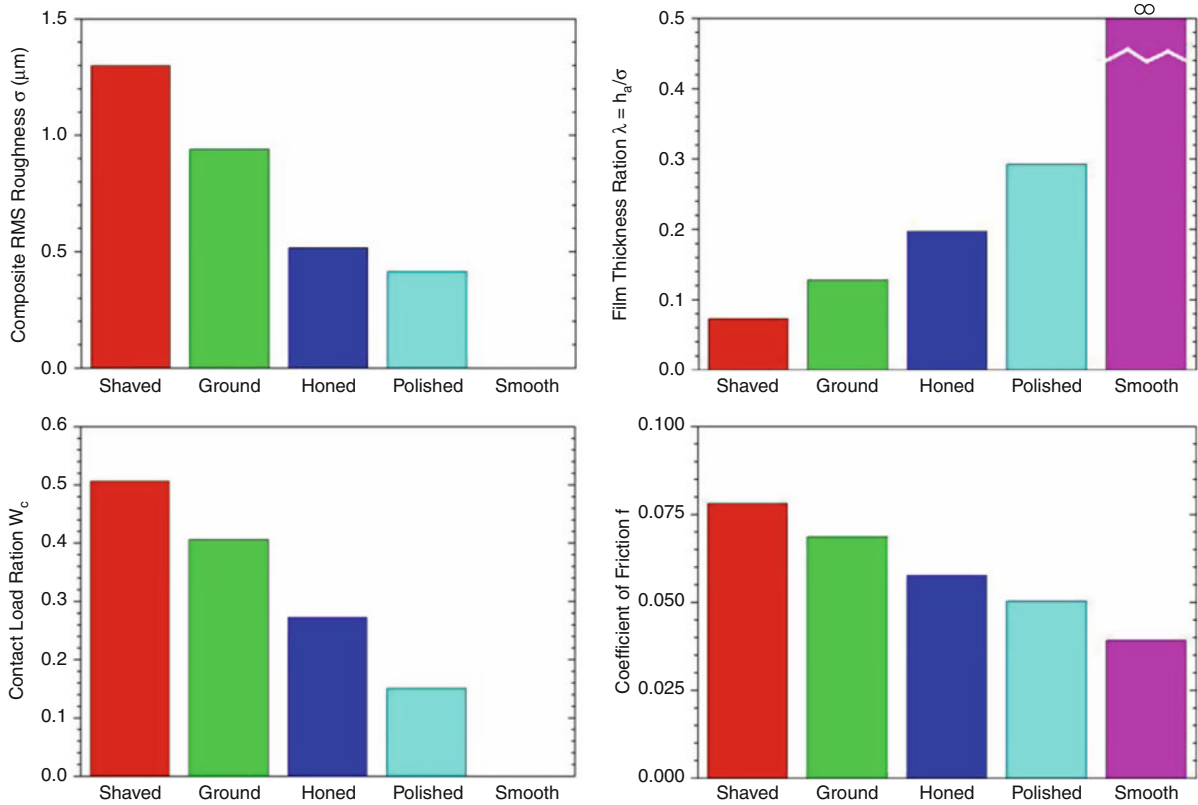
Sample Mixed EHL Results

Figure 3 shows an example case for a spur gear set (a pinion of 11 teeth vs. a gear of 45 teeth). The mixed EHL analysis is conducted for the critical contact location of LPSTC (lowest point of single tooth contact) below pitch line, where severe pitting failure and/or excessive wear are often found. Digitized machined surfaces given in Fig. 2 are used and contour graphs for resulting EHL pressure and film thickness distributions are illustrated in the figure. For comparison purpose, the solution from ideally smooth surfaces is also given on the right-hand side. It can be seen that when the surfaces are smooth, there is no



Mixed EHL in Gears, Fig. 2 Sample gear tooth surfaces

Mixed EHL in Gears, Fig. 3 Sample results from the 3D line contact mixed EHL model for a spur gear set (11/45) under LPSTC conditions ($P_H = 2.919 \text{ GPa}$, $U^* = 0.134 \times 10^{-10}$, $W^* = 1.037 \times 10^{-3}$, $G^* = 5494$, $S = 114.3\%$)



Mixed EHL in Gears, Fig. 4 Effect of gear tooth surface finish on mixed EHL characteristics

asperity contact and the full-film EHL of central film thickness about $h_c = 100$ nm is observed. The pressure distribution is close to that of the Hertzian contact. If the surfaces are rough, however, significant asperity contacts take place, illustrated in the film thickness contours by light yellow areas. The pressure distribution becomes peaky and high spikes may exceed 2^*P_h as indicated by light yellow colored areas in the pressure contours. Note that the pressure and film thickness distributions appear to be well correlated with the machined surface topography.

Figure 4 presents a summary of mixed EHL results for the same set of solutions. The film thickness ratio, also called λ ratio, is defined as the average film thickness divided by the composite RMS roughness, $\lambda = h_0/\sigma$, and the contact load ratio, W_c , is the load supported by the rough surface asperity contacts divided by the total load. The λ ratio is an important parameter representing lubrication effectiveness, while the contact load ratio indicates contact severity in mixed lubrication. These parameters are useful in design and failure analysis. The cases presented here are typical for heavy-duty gears. It is

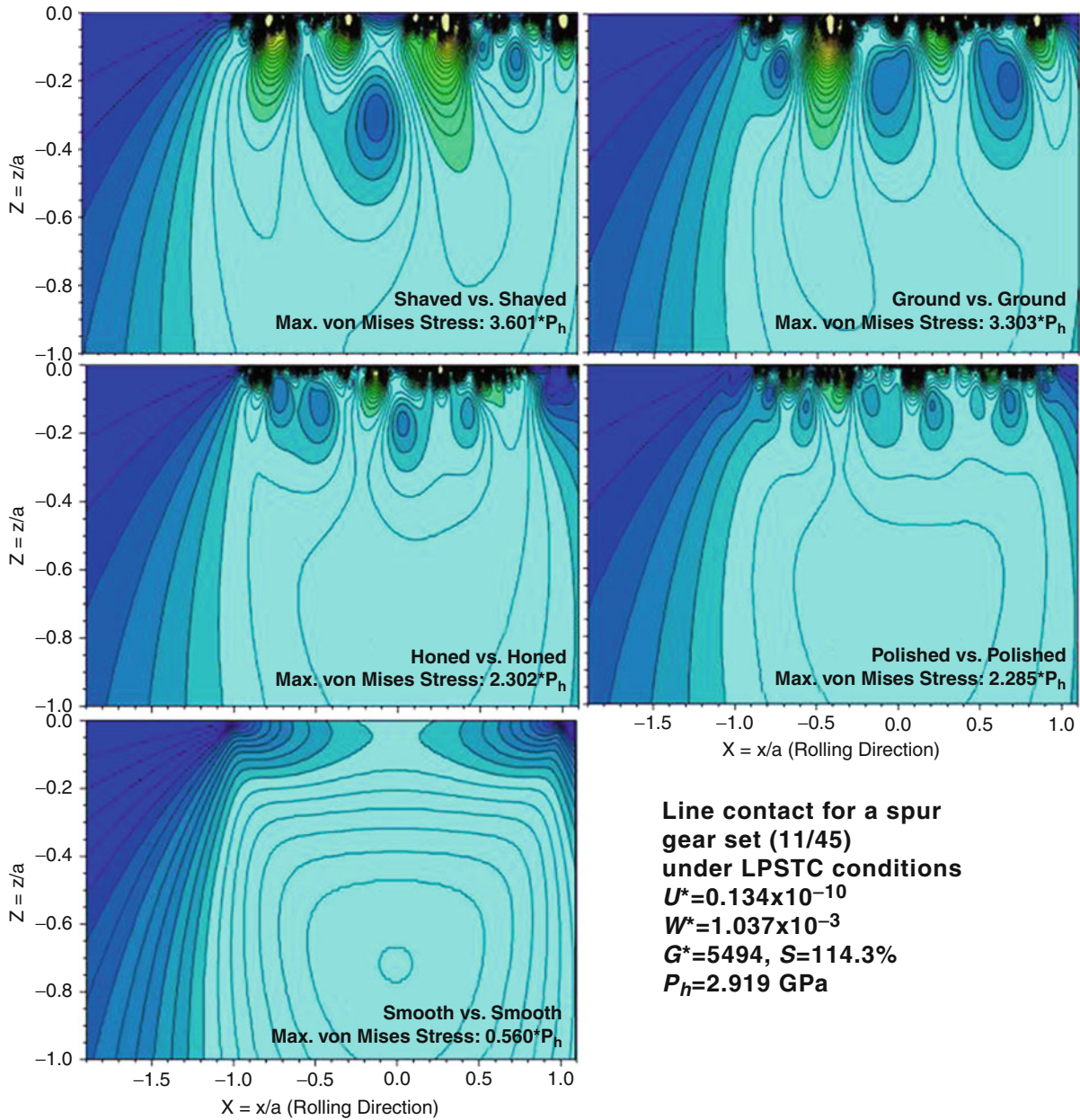
obvious that honed and polished tooth surfaces with smaller roughness provide significantly better lubrication performance and less severe asperity contacts leading to much reduced risk of failure, compared with very rough surfaces such as the shaved.

Subsurface Stresses

Figure 5 presents subsurface von Mises stress fields resulting from mixed EHL pressure distributions shown in Fig. 3. It appears that severe asperity contacts lead to a significant increase in maximum von Mises stress and movement of its location towards the surface. This may cause early surface failures such as excessive wear and pitting due to contact fatigue.

Mixed EHL Friction in Gears

Friction in gears is of great importance in practice, as it relates directly to efficiency and thermal behavior. Although the efficiency of gearboxes is typically high, further improvement is required by today's energy-saving and environmental protection strategies, and even a small amount of frictional loss may cause considerable



Mixed EHL in Gears, Fig. 5 Effect of gear tooth surface finish on subsurface von Mises stress field

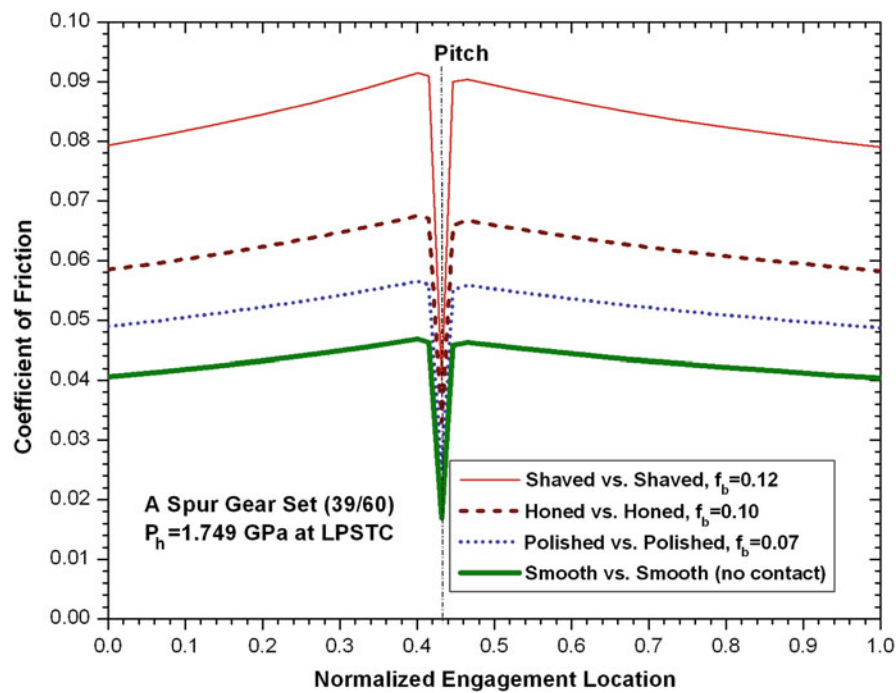
temperature increase, which may in turn lead to a significant reduction of viscosity and may negatively affect lubrication performance, resulting in further increase in friction and possible lubrication breakdown.

Friction in mixed EHL consists of two parts. In hydrodynamically lubricated areas the friction is generated from lubricant shearing in the EHL films. It is usually low and can be calculated by using a non-Newtonian

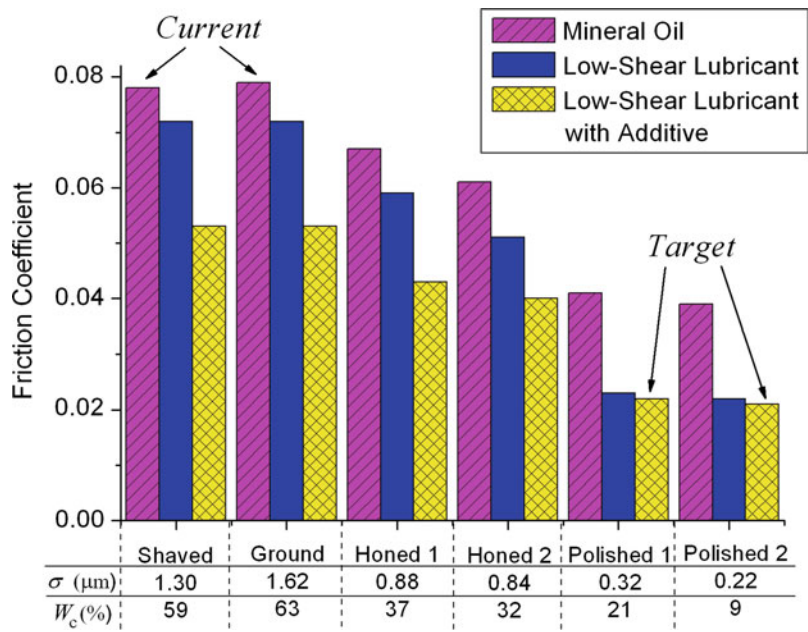
viscous-elastic fluid model, such as the one proposed by Bair and Winer (1978) given as follows:

$$\dot{\gamma} = \frac{\dot{\tau}}{G_x} - \frac{\tau_L}{\eta} \ln \left(1 - \frac{\tau}{\tau_L} \right)$$

where τ_L and G_x are the limiting shear stress and limiting elastic modulus, respectively, which are properties of the lubricant and also functions of pressure and temperature.



Mixed EHL in Gears, Fig. 6 Sample results of gear tooth contact friction



Mixed EHL in Gears, Fig. 7 Effects of surface finish, lubricant and boundary additive on friction

Generally, advanced synthetic gear lubricants exhibit low friction behavior because of their low limiting shear stress.

In asperity contact areas, friction is usually much higher, determined by boundary lubrication properties. Boundary friction is difficult to model due to complicated interfacial physics and chemistry, but it is easy to estimate experimentally. Boundary friction coefficient f_B is often in a range of 0.07–0.15 for lubrication systems with commonly used steels and industrial lubricants. If f_B is provided as an input, the total friction in the mixed EHL can be predicted, as described in detail by Martini et al. (2007). It is obvious that the total friction is greatly affected by how much asperity contact exists in the mixed EHL system. Generally, the larger the contact load ratio W_c , the greater the friction.

Figure 6 presents tooth contact friction variations during a meshing cycle in a spur gear set. Note that at pitch point the sliding friction is supposed to be zero or close to zero due to the pure rolling condition. Figure 7 is reproduced from Martini et al. (2007) showing practical friction reduction means for gears. It can be seen that the friction can be significantly reduced by the following three major means, which are the technologies currently available:

1. Use better finished surfaces, such as honed and polished, with considerably lower RMS roughness. In this way the surface asperity contacts can be minimized and lubrication improved.
2. Choose advanced lubes, such as synthetic gear oils, with low limiting shear stress, so that the hydrodynamic component of friction due to lubricant shearing can be reduced.
3. Reduce boundary friction coefficient by employing advanced boundary additives. This is very effective, especially for heavy-duty gears in which asperity contacts are severe.

It can be seen that by using these three techniques combined, the total tooth contact friction can be reduced significantly, say by up to 70%, from the situation of shaved surface finish with regular mineral oil and no advanced boundary additive (which widely exists currently in practice) to improved condition with honed or polished surfaces and low-shear synthetic lubes plus advanced additives.

Key Application

Gears are key components that are widely used in various vehicles and machineries for transmitting power and motion through tooth surface contacts. Most gears operate in the mixed EHL regime, in which both the lubricant

films and surface asperity contacts coexist and neither can be ignored. A good understanding of mixed EHL characteristics is vital to improvements of gear performance, efficiency, and durability as well as design optimization and failure prevention. Available mixed EHL models can be used to predict pressure and film thickness distributions, λ ratio, asperity contact severity, subsurface stresses, and friction. These parameters are extremely important in gear research, design, and problem solving. Based on the understanding of mixed EHL in gears, practical means for gear friction reduction are recommended. With these technologies currently available, gear tooth contact friction can be greatly reduced and efficiency improved. Effective failure prevention also requires the deep understanding of mixed EHL, but its detailed discussion is beyond the scope of this essay.

Nomenclature

a	Half-width of Hertzian contact zone
E'	Effective elastic modulus
F	Tooth load
f_B	Boundary friction coefficient
$G^* = \alpha E'$	Dimensionless material parameter
h_a	Average film thickness
l_e	Effective length of contact line
P_h	Maximum Hertzian pressure
R_a	Mean roughness height
R_q	RMS roughness
$S = V/U$	Slide-to-roll ratio
U	Rolling velocity (or entraining velocity)
$U^* = \eta_0 U / (E' R_x)$	Dimensionless speed parameter
V	Sliding velocity
$W^* = F / (E' R_x l_e)$	Dimensionless load parameter
W_c	Contact load ratio (load supported by asperity contacts divided by total load)
α	Pressure-viscosity exponent used in pressure-viscosity equation $\eta = \eta_0 \text{EXP}(\alpha p)$
η, η_0	Viscosity and viscosity under ambient condition, respectively
$\lambda = h_a / \sigma$	Film thickness ratio
$\sigma = (R_{q1}^2 + R_{q2}^2)^{0.5}$	Composite RMS roughness

Cross-References

- [3D Line Contact EHL](#)
- [Deterministic Models of Rough Surface EHL](#)

- [EHL Governing Equations](#)
- [Elastohydrodynamic Lubrication \(EHL\)](#)
- [Mixed EHL](#)
- [Sliding Wear in Mixed EHL Contacts](#)

References

- S. Bair, W.O. Winer, Rheological response of lubricants in EHD contacts, in *Proceedings of the 5th Leeds-Lyon Symposium on Tribology* (The University of Leeds, England, 1978), pp.162–169
- D. Dowson, G.R. Higginson, New roller bearing lubrication formula. *Engineering* **192**, 158–159 (1961)
- D. Dowson, G.R. Higginson, *Elastohydrodynamic Lubrication* (Pergamon, London, 1966)
- A. Gu, Elastohydrodynamic lubrication of involute gears. *J. Eng. Industry* **95**, 1164–1170 (1973)
- M.J.A. Holmes, H.P. Evans, R.W. Snidle, Analysis of mixed lubrication effects in simulated gear tooth contacts. *J. Tribol.* **127**, 61–69 (2005)
- Y.Z. Hu, D. Zhu, A full numerical solution to the mixed lubrication in point contacts. *J. Tribol.* **122**, 1–9 (2000)
- H.M. Martin, Lubrication of gear teeth. *Engineering* **102**, 109–121 (1916), London
- A. Martini, D. Zhu, Q. Wang, Friction reduction in mixed lubrication. *Tribol. Lett.* **28**, 171–181 (2007)
- N. Ren, D. Zhu, W. Chen, Y. Liu, Q. Wang, A three-dimensional deterministic model for rough surface line contact EHL problems. *ASME J. Tribol.* **131**(011501), 1–9 (2009)
- K.L. Wang, H.S. Cheng, A numerical solution to the dynamic load, film thickness, and surface temperatures in spur gears, part I: analysis. *J. Machine Design* **103**, 177 (1981)
- D. Zhu, A comprehensive analysis for mixed elastohydrodynamic lubrication performance and friction in spur and helical gears, in *Proceedings of the ASME/STLE 2009 International Joint Tribology Conference*, Park Ridge, Illinois, USA, IJTC2009-15148 (2009)
- D. Zhu, H.S. Cheng, A comprehensive analysis for contact geometry, kinematics, lubrication performance, bulk and flash temperatures in helical gears, in *Proceedings of the 17th Leeds-Lyon Symposium on Tribology* (The University of Leeds, England, 1991), pp.383–389
- D. Zhu, Y.Z. Hu, The study of transition from full film elastohydrodynamic to mixed and boundary lubrication, in *The Advancing Frontier of Engineering Tribology, Proceedings of the 1999 STLE/ASME H.S. Cheng Tribology Surveillance*, Park Ridge, Illinois, USA, (1999), pp.150–156
- D. Zhu, A. Martini, W. Wang, Y.H. Hu, B. Lisowski, Q. Wang, Simulation of sliding wear in mixed lubrication. *ASME J. Tribol.* **129**, 544–552 (2007)
- D. Zhu, N. Ren, Q. Wang, Pitting life prediction based on a 3-D line contact mixed EHL analysis and subsurface von Mises stress calculation, *ASME J. Tribol.* **131**, 041501 1–8.

Mixed Elastohydrodynamic Lubrication

- [Mixed EHL](#)

Mixed Lubrication

MIHAI B. DOBRICA¹, MICHEL FILLON²

¹Tecnitas, Bureau Veritas Group, Levallois-Perret Cedex, France

²Institut Pprime, Department GMSC, CNRS - Université de Poitiers - ENSMA, Futuroscope Chasseneuil Cedex, France

Synonyms

[Lubrication state between boundary and hydrodynamic regimes](#)

Definition

Mixed lubrication is an operating state (regime) of a lubricated contact in which surface roughness (► [Surface Roughness](#)) significantly affects the performance of the contact. It may occur with conformal contact lubrication, such as journal bearing lubrication. Traditionally, as suggested by the term *mixed*, it is thought that both hydrodynamic lubrication and asperity contact have to be present for a contact to be considered as functioning in the mixed regime (Fig. 1, II). However, between the theoretical limit of the mixed lubrication regime and the full hydrodynamic regime (with film thickness much higher than roughness, [► [Hydrodynamic Lubrication](#)]), there is a region where surface roughness still has significant influence on contact performance, even if asperity contact is not present.

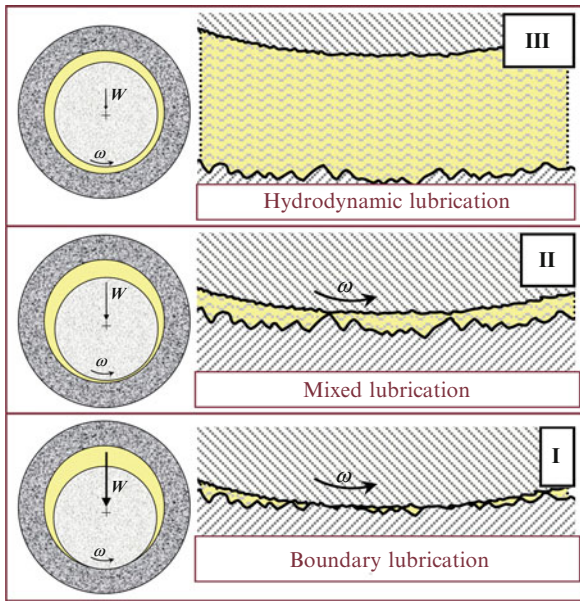
In the absence of prior knowledge about roughness influence, it can be considered that a lubricated contact is in the mixed regime when the overall separation between the contacting bodies is of the same order of magnitude as, or smaller than, the surface roughness. However, precise limits of the mixed lubrication regime are yet to be defined.

The mixed lubrication (Fig. 1) regime is situated between full hydrodynamic lubrication (film thickness is significantly higher than surface roughness) and boundary lubrication (solid-to-solid contact becomes dominant; lubrication has little influence on the contact behavior). These three lubrication regimes are generally presented on the Stribeck curve (Fig. 2).

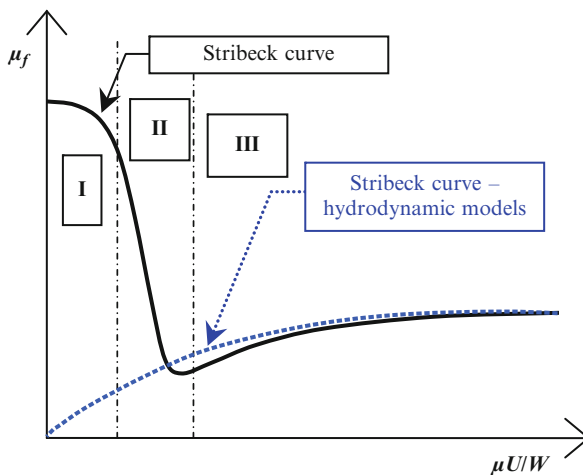
Scientific Fundamentals

Effects of Operation in Mixed Lubrication

While theoretically the minimum friction coefficient of a lubricated contact is obtained at the junction between



Mixed Lubrication, Fig. 1 Hydrodynamic, mixed, and boundary lubrication in a journal bearing



Mixed Lubrication, Fig. 2 Stribeck curve

full-film and the mixed lubrication regimes, this state is not usually sought after, or desirable, in general machinery. Mixed lubrication produces significant oscillations in local pressures and can lead to important stress peaks, which can have adverse effects on both solids and lubricant. Additionally, a lubricant film thickness comparable to the surface roughness is generally too small to allow sufficient lubricant flow and excessive heating may become a problem.

In moderate to severe mixed lubrication configurations, asperity contact is present and becomes increasingly severe. These configurations pose the threat of extensive localized wear that can eventually lead to failure (wiping of the bearing surface). To these one may add the poor understanding of mixed lubrication-related phenomena, and it becomes clear why this lubrication regime is less desirable. It should be noted that mixed lubrication is often unavoidable at the start and stop of rotating machinery.

Occurrence

Extensive functioning of lubricated conformal contacts in the mixed lubrication regime should be avoided. In fluid film bearings, it may still occur for short periods of time or during unplanned events. One obvious situation where mixed lubrication occurs is during startup and shutdown of fluid film bearings that are not equipped with expensive hydrostatic lift devices. Mixed lubrication may also occur, at least at the ends of reciprocating motion of a piston, in an internal combustion engine. The regime is also encountered in rotating machinery that is run at very low speeds or excessively high loads. Misalignment is another factor leading to a localized mixed lubrication regime in bearings. Occasional severe dynamic loading can also lead, for brief amounts of time, to very low film thicknesses and, hence, mixed lubrication.

In mechanical face seals (► [Mechanical Seals](#)), the film thickness is much smaller than in hydrodynamic bearings (in order to prevent leakage), but so is the roughness of the mating surfaces. Consequently, mixed lubrication is not more likely to develop in normal functioning conditions. Similar mechanisms (start-up and shut-down, misalignment) can lead to mixed lubrication in mechanical face seals. Additionally, surface waviness is often responsible for mixed lubrication and localized wear in these components. In hydraulic seals, and particularly in reciprocating seals, the mixed lubrication regime is more often encountered. Here, low leakage is privileged at the expense of a fully lubricated interface. The increased compliance of these seals allows for reduced asperity contact and peak pressures (and, hence, reduced wear) despite film thicknesses comparable to the asperity height.

Mixed lubrication is unavoidable in lubricated counterformal contact interfaces, such as teeth of gears, rolling and ball bearings.

Theoretical Aspects and Related Phenomena

There are two main phenomena encountered in the mixed lubrication of surfaces. The first is inter-asperity lubrication (thin film lubrication). From a theoretical point of

view, as long as the film thickness is sufficiently large with respect to the size of the molecules of lubricant, classical lubrication theory remains valid. However, this theory is to be applied at a micrometric (or sub-micrometric) scale in order to account for the local variations in film thickness induced by the roughness of the mating surfaces. The second phenomenon encountered in the more severe mixed lubricated contacts is inter-asperity (or rough) contact. Numerous studies have focused on inter-asperity contact modeling, and the theory is satisfactory for steady-state rough contacts and for transient configurations where there is no slip at the interface between the contacting bodies. Further theoretical developments are required in order to account for the relative motion of the contacting bodies at the roughness scale. If the mixed lubrication regime is mild and only limited and spaced asperity contact is present, Hertz theory can be applied for each contacting asperity, as a first, coarse approximation.

One particular phenomenon that has yet to be studied is the transition from lubrication to contact. While continuum mechanics theories apply in both lubricated and contact areas, they no longer hold in the transition regions where the film thickness is of the same order of magnitude as the size of lubricant molecules. According to continuum laws, lubricant pressure in an increasingly loaded contact will increase exponentially as the film thickness decreases, up to the point where solids would plastically deform in order to make room for the lubricant. Consequently, in theory, direct solid-to-solid contact cannot develop if the contact that was initially fully lubricated (i.e., a fluid film will always exist between the two mating bodies). The extent to which lubricant molecules subsist in the contacting regions is debatable, but experimental observations show that the interface behaves completely different with respect to full film lubricated contacts (solid wear and debris are present and the electrical resistance at the interface is reduced).

Most general lubrication-related phenomena are equally encountered in the mixed state, although they develop at a significantly smaller scale. Piezo-viscosity, for example, is likely to be important in mixed lubricated contacts where, at least locally, large pressures peaks are present. Cavitation ([► Cavitation Phenomena and Numerical Analysis](#)), called “micro-cavitation” when it develops at the roughness scale, may appear in the micro-pits generated by roughness, especially in the regions where sufficient lubricant flow is prevented by asperity contact. Thermal effects are also present, and large local temperature peaks have been reported in the asperity contact regions ([► Flash Temperature in EHL Contacts](#)). As for the contact regions, since wear and

surface roughness modification is usually associated with mixed lubrication, it is likely that plastic asperity deformation is present in the more severe mixed states.

Elastic deformations of solids (and eventually plastic) are present and generally important in mixed-lubricated contacts, even in the absence of severe asperity contact. Therefore, deformation of solids cannot be ignored, unless the mixed lubrication state is caused by very low sliding velocities in the absence of significant loads.

Prediction of Mixed Lubrication

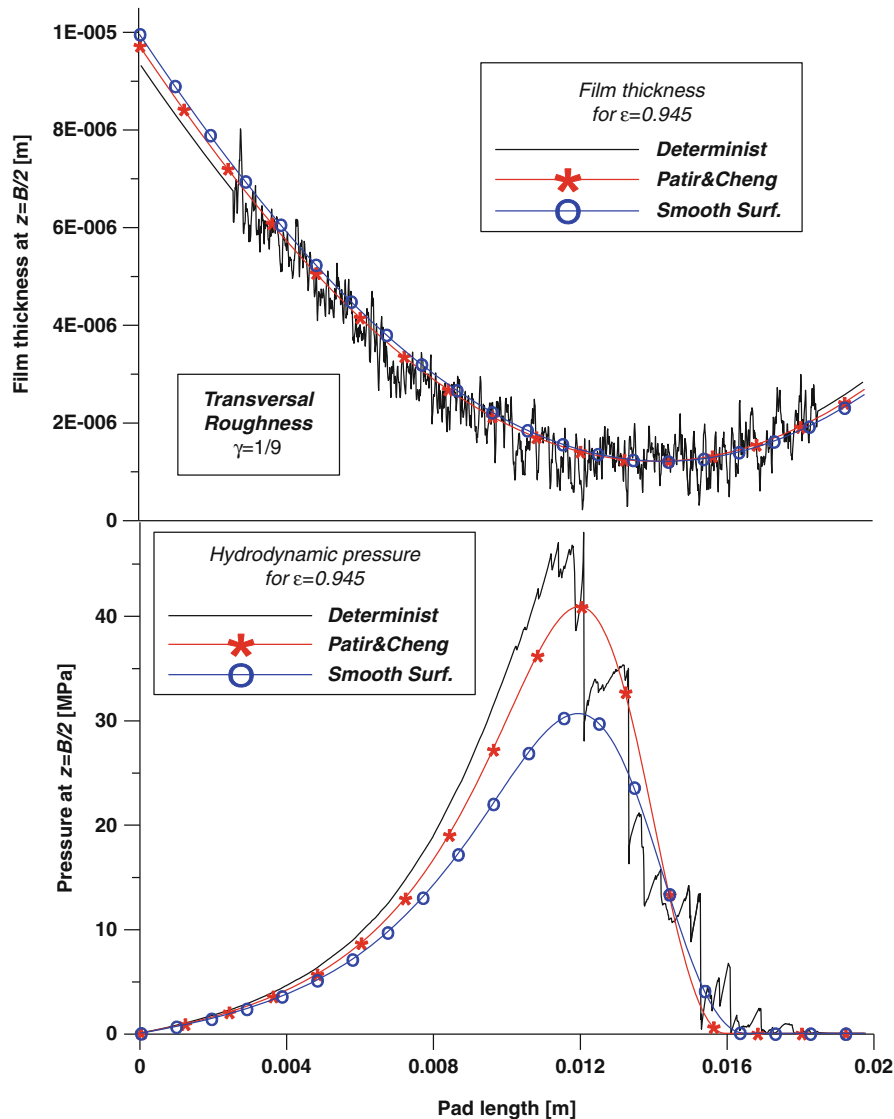
Prediction of the performance of contacts functioning in the mixed lubrication regime is generally obtained through numerical modeling. Due to the micrometric scale at which mixed lubrication takes place, local experimentation is particularly difficult. Even if measurements could be taken at the roughness level in a lubricated contact functioning in the mixed regime, it is likely that local oscillation present in both pressure and film thickness measurements would make interpretation of these results difficult. The works of Höhn et al. (2006) are among the few studies that provide experimental results for rolling contacts functioning in mixed lubrication and having realistic grinded surfaces (as opposed to surfaces presenting a single geometric feature). Their conclusions indicate a better performance of surfaces with transverse grinding as opposed to circumferential grinding, and an increase in pressure oscillations (and hence peak pressure) with the increase in surface roughness. These observations were later confirmed by several numerical models.

The numerical modeling of mixed lubrication contacts is discussed in detail in the next section.

Numerical Modeling

A large majority of the research related to the mixed lubrication regime focuses on numerical modeling. Although most of the existing models have been predominantly applied in counterformal contacts (where the mixed lubrication regime is more severe), no distinction will be made here inasmuch as similar theoretical and numerical concerns are encountered when modeling conformal contacts.

There are two main categories of theoretical approaches used when describing the mixed lubrication regime: stochastic models and deterministic models ([Fig. 3](#)). Deterministic models actually divide the domain into extremely small elements, small enough to catch the influence of every single asperity. Classical lubrication and contact theory is subsequently applied on this very fine mesh, thus providing information on what happens at the roughness level. Contrary to deterministic approaches,



Mixed Lubrication, Fig. 3 Deterministic, stochastic, and hydrodynamic film thickness and pressure in a mixed-lubricated journal bearing (Dobrica et al. 2006)

stochastic models were developed for larger contact domains, where roughness-level meshing is difficult. Stochastic approaches statistically transpose the roughness influence recorded at a micrometric scale (using deterministic models or single asperity models) up to a larger scale, where each mesh element actually contains a large number of asperities. While local information is obviously lost in the latter process, such models do take into account some of the roughness influence.

Stochastic modeling. Stochastic modeling is the approach most often adopted when studying conformal

contacts functioning in the mixed lubrication regime. These models are relatively simple and easy to integrate in existing numerical algorithms, and, as stated before, are designed for coarse meshes that allow for fast simulations. Furthermore, stochastic models avoid the difficult theoretical issue of the transition from lubrication to contact. A stochastic model for mixed lubricated contacts generally contains two sub-models: a stochastic thin film lubrication model (hydrodynamic component) and an asperity contact stochastic model (contact component).

The most used stochastic thin film lubrication model is the one proposed by Patir and Cheng (1978). It led to the so-called “average Reynolds equation” (► [Average Reynolds Equations](#)):

$$\frac{\partial}{\partial x} \left(\phi_x \frac{h^3}{12\mu} \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left(\phi_y \frac{h^3}{12\mu} \frac{\partial p}{\partial y} \right) = \frac{U_1 + U_2}{2} \frac{\partial h_T}{\partial x} + \frac{U_1 - U_2}{2} \sigma_s \frac{\partial \phi_s}{\partial x} + \frac{\partial h_T}{\partial t}$$

in which the roughness influence (σ_s = R.M.S. roughness) is accounted for using two pressure flow factors (ϕ_x and ϕ_y) and a shear flow factor (ϕ_s). This equation describes the fluid flow between two rough surfaces separated by a thin fluid film, in relative motion with respect to each other. Values of the flow factors were determined by Patir and Cheng (1979) through deterministic simulations of small sliding parallel contacts having rough surfaces. It should be noted that the values provided by Patir and Cheng had an exemplifying purpose, and were accurate only for specific types of surface roughness (longitudinal/isotropic/transversal). They recommended that adequate flow factors be computed (through deterministic simulations) for each specific contact in which the model would be applied. A more advanced stochastic model accounting for micro- and macro-cavitation was proposed by Harp and Salant in 2001. To date, this model has been only rarely applied.

The first stochastic asperity contact model, often employed in mixed-lubrication related studies, is the one proposed by Greenwood and Williamson (1966, ► [Contact of Rough Surfaces: The Greenwood and Williamson/Tripp, Fuller and Tabor Theories](#)). It is based on the Hertzian contact solution between an elastic spherical asperity and a flat rigid body. In function of the surface roughness (σ), mean asperity radius (R), asperity density (η), mechanical properties (E), nominal contact area (A_n), and the average separation between the two contacting bodies (d), this model estimates the average contact pressure (P_c) and contact area (A_c):

$$A_c(d) = \frac{\pi \eta A_n R}{\sigma \sqrt{2\pi}} \int_d^\infty (z - d) e^{\frac{1}{2} \left(\frac{z}{\sigma} \right)^2} dz$$

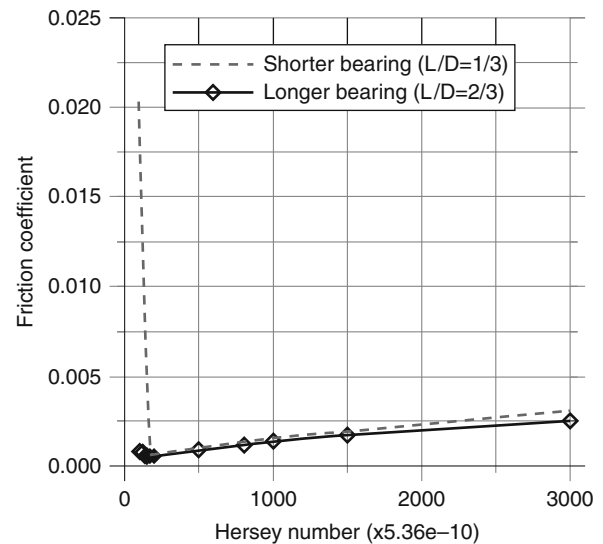
$$P_c(d) = \frac{\frac{4}{3} \eta A_n E R^{\frac{1}{2}}}{\sigma \sqrt{2\pi}} \int_d^\infty (z - d)^{\frac{3}{2}} e^{\frac{1}{2} \left(\frac{z}{\sigma} \right)^2} dz$$

Several stochastic asperity-contact models have been proposed since and can take into account plastic deformations, elliptical asperity shape, asperity interaction, and so on (see [► [Stochastic Contact Theories: Theories of](#)

[Surface Roughness and Applications to Contact Mechanics](#)] for more details). Stochastic contact models can also be derived through deterministic simulation (Lee and Ren 1996) in a way that is similar to the approach used for determining flow factors for the thin film lubrication models. The contact between two bodies having representative rough surfaces is deterministically modeled for various levels of average clearance, and the results for the average pressure and contact area are recorded. Analytical relations fitting these results are then determined and used in stochastic modeling.

A complete numerical study based on a stochastic mixed-lubrication model was discussed by Wang et al. (2006). The authors used the (Patir and Cheng 1978) thin film lubrication model, combined with the (Lee and Ren 1996) stochastic contact approach, in order to derive Stribeck curves for journal bearings functioning in the mixed lubrication regime. Figure 4 (Wang et al. 2006) presents the evolution of the friction coefficient in function of the Hersey number (Stribeck curve), as obtained by stochastic modeling.

The main disadvantages of stochastic models are the current lack of validation and the restrictive hypotheses used in their derivation (particular roughness type or periodicity, particular asperity shape, etc.). Stochastic approaches remain the only current option for modeling large mixed-lubricated conformal contacts, since roughness level meshing is impractical for such applications. Stochastic models, especially when using factors based



Mixed Lubrication, Fig. 4 Stribeck curve for hydrodynamic journal bearings (Wang et al. 2006)

on deterministic results obtained for small sample contacts with representative surface roughness, have the potential to provide reasonably accurate results in a fraction of the time required for deterministic modeling.

Deterministic modeling. A literature review shows that deterministic models for lubricated conformal contacts functioning in the mixed regime are scarce. Due to roughness level meshing, these models are particularly demanding in terms of computing time, and they require special convergence acceleration techniques in order to be practical even in the smallest EHL contacts. The most time-consuming component of these models is the displacement calculus that, in the absence of particular acceleration methods, is a N^2 process (i.e., on a mesh having 10^6 nodes, each displacement calculus requires 10^{12} multiplications – largely impractical even on modern computers). The contact component of deterministic models is also time consuming, since displacement due to contact pressure has to be computed several times before convergence is acquired. In order to speed up the displacement computation time, the multi-level multi-integration technique (Brandt and Lubrecht 1990) or FFT and DC-FFT models (Ju and Farris 1996; Liu et al. 2000, FFT-based methods for contact mechanics) can be used. An evaluation of some of these techniques for rough contact simulation was given by Allwood (2005). The solving of Reynolds' equation can be accelerated by the use of multi-grid methods (► [Multi-Grid Method](#)).

Currently, there are at least three deterministic models that, according to their respective authors, take into account both thin film lubrication and rough surface contact, coupled through an elastic displacement calculus. They all make some simplifying hypotheses in order to avoid the incertitude related to the transition between lubricated and contact regions. Only one of these models is applied to conformal contacts.

The first completely deterministic mixed lubrication model, which takes into account both thin film lubrication and rough contact, solving each fluid-asperity boundary, has been proposed by Hua et al. (1997). The Reynolds equation, discretized using the finite differences method and solved with the multigrid technique, is used for flow modeling. The theoretical contact model is that between a smooth rigid sphere and a rough semi-infinite elastic body, which provides analytical values for the influence coefficients used in displacement calculus. In exchange, the model is directly applicable only to EHL point contacts. Mechanical deformations are computed using a FFT-based algorithm, while contact pressure in the penetrating nodes is determined using FFT deconvolution (IFFT). This yields a fast numerical algorithm, applied

for point EHL contacts between a rough surface and a smooth sphere on meshes having up to 421×421 nodes. There are several factors that make contact possible in this model. Firstly, contact is privileged over hydrodynamic lubrication during the iterative process. At each iteration, the contact solution is first computed for the penetrating nodes, and then the hydrodynamic pressure is computed for the rest of the domain. Secondly, pressure is cut off at the plasticity limit of the softer material. Consequently, contact is guaranteed to develop in the plastic regions (where further local displacement is prevented due to the capped pressure level). Finally, mesh convergence tests were not conducted, and it is possible that an insufficiently refined mesh was used. In this case, local hydrodynamic pressure peaks in the transition regions (from lubrication to contact) may be under-evaluated, and so would be the corresponding displacement. This issue is discussed in detail by Venner (2005).

The second deterministic lubrication model, widely employed by its authors in various studies of mixed lubricated counter-formal contacts, was first presented in Zhu and Hu (1999). According to the authors, the Reynolds equation is applicable in both lubricated and contact regions ($h \leq h_{\text{lim}}$),

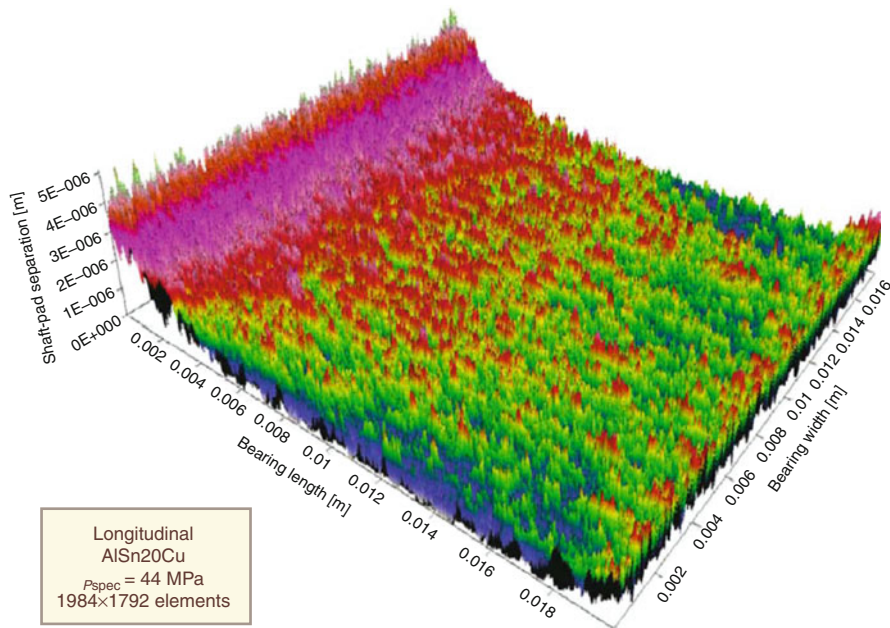
$$\begin{aligned} & \frac{\partial}{\partial x} \left(\frac{\rho}{12\mu} h^3 \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{\rho}{12\mu} h^3 \frac{\partial p}{\partial y} \right) \\ &= U \frac{\partial(\rho h)}{\partial x} + \frac{\partial(\rho h)}{\partial t} \quad \text{for } h > h_{\text{lim}} \end{aligned}$$

$$U \frac{\partial h}{\partial x} + \frac{\partial h}{\partial t} = 0 \quad \text{for } h \leq h_{\text{lim}}$$

the pressure terms being removed from the latter (Reynolds reduced equation, RRE). A very small threshold film thickness value (h_{lim}) is used for switching between contact and lubricated domains. According to the authors, the pressure is updated in the contact regions, even if the pressure term is missing from the reduced equation ($h \leq h_{\text{lim}}$) (Zhu and Hu 1999).

An interesting debate took place between Venner (2005) and Zhu (2007) about this model and provides further insight into common problems related to mixed-lubrication simulation. Issues concerning all deterministic models, like mesh size and grid convergence, transition from lubrication to contact, or the use of cut-off values for film thickness are discussed in detail.

The model has been subsequently improved by the progressive mesh densification method (PMD) to ensure a reasonable and efficient mesh-converged solution with satisfactory numerical accuracy. In order to accelerate the



Mixed Lubrication, Fig. 5 Shaft-pad separation (film thickness) in a mixed lubricated journal bearing (elastic and plastic deformations are taken into account) (Dobrica et al. 2008)

surface deformation calculation, which is a dominant part of the total solution process, the discrete convolution-fast Fourier transform (DC-FFT) method (Liu et al. 2000) is incorporated into the EHL model. Model validation was conducted in a series of studies of thin-film fully lubricated contacts (sphere-plane contacts, configurations including simple dimples) that show good agreement between numerical predictions and experimental data.

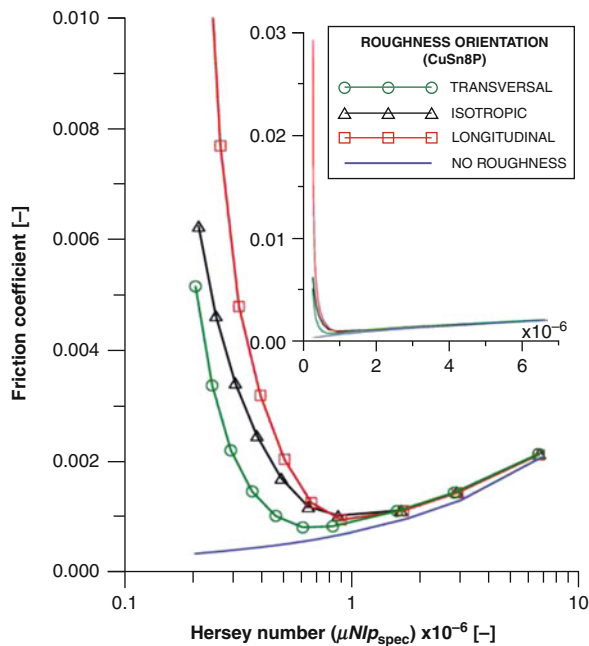
A third deterministic model for mixed lubrication was presented by Dobrica et al. (2006), and was further extended in Dobrica et al. (2008). This is the first deterministic model to be applied to small conformal contacts—partial journal bearings having a diameter of 3 cm, modeled using a mesh having 3.6×10^6 elements (Fig. 5). The lubricant flow is modeled with the Reynolds equation, made discrete using the finite volume method and solved using a SOR solver. Displacement and contact solution required specific modeling techniques, due to the particular geometry of journal bearings, which does not accept analytical solutions for the influence coefficients. Hence, influence coefficients were computed through finite element simulations, using a two grid approach. The same two-grid approach was used in order to accelerate displacement computations. Contact pressure was computed using an active-set/contact-set approach (Allwood 2005), with a Gauss-Seidel solver.

This model also made a forced hypothesis in order to allow the presence of rough contact: lubricant pressure is cut off at the plasticity limit of the softer material, and it is assumed that only contact can produce plastic deformations.

Stribeck curves (Fig. 6) were traced for the considered journal bearings, considering three types of surface roughness (longitudinal/isotropic/transversal), for two types of bearing pad alloys. It was concluded that transversal roughness offers the best performance in mixed-lubricated contacts, and that roughness effects are more pronounced in more rigid contacts. Comparisons between deterministic and stochastic results (Patir and Cheng) were made for the hydrodynamic component of the model. It was concluded that Patir and Cheng offer an acceptable approximation to roughness effects in lubrication, which could probably be improved by computing surface-specific flow factors.

Key Applications

Operation of conformal lubricated contacts in the mixed lubrication regime is generally avoided, and there are very few actual “applications” where this lubrication regime is intended for normal operation. Crankshaft bearings in Formula One engines are one example where mixed lubrication is used to minimize friction. However, these devices



Mixed Lubrication, Fig. 6 Stribeck curves for a small partial journal bearing having transversal/isotropic/longitudinal roughness on a bronze pad (Dobrica et al. 2008)

are expected to run only for limited amounts of time, and extensive wear prediction studies are conducted to determine their lifespan. Reciprocating hydraulic seals are another application designed to operate in mixed lubrication, minimizing leakage at the expense of some additional wear.

Severe mixed lubrication is common and unavoidable in lubricated counterformal contact interfaces, such as teeth of gears, rolling and ball bearings, and camshaft-follower contacts. For information on the mixed lubrication of counterformal contacts.

Cross-References

- Average Reynolds Equations
- Cavitation Phenomena and Numerical Analysis
- Contact of Rough Surfaces: The Greenwood and Williamson/Tripp, Fuller and Tabor Theories
- FFT-Based Methods for Contact Mechanics
- Flash Temperature in EHL Contacts
- Hydrodynamic Lubrication
- Mechanical Seals
- Multi-grid Method
- Stochastic Contact Theories: Theories of Surface Roughness and Applications to Contact Mechanics
- Surface Roughness

References

- J. Allwood, Survey and performance assessment of solution methods for elastic rough contact problems. *ASME J. Tribol.* **127**, 10–23 (2005)
- A. Brandt, A.A. Lubrecht, Multilevel matrix multiplication and fast solution of integral equations. *J. Comp. Phys.* **90**, 348–370 (1990)
- M.B. Dobrica, M. Fillon, P. Maspeyrot, Mixed elastohydrodynamic lubrication in partial journal bearings – comparison between deterministic and stochastic models. *ASME J. Tribol.* **128**, 778–788 (2006)
- M.B. Dobrica, M. Fillon, P. Maspeyrot, Influence of mixed-lubrication and rough elastic-plastic contact on the performance of small fluid film bearings. *STLE Tribol. Trans.* **51**, 699–717 (2008)
- J.A. Greenwood, J.B.P. Williamson, Contact of nominally flat surfaces. *Proc. Roy. Soc.* **A295**, 300–319 (1966)
- B.-R. Höhn, K. Michaelis, O. Kreil, Influence of surface roughness on pressure distribution and film thickness in EHL contacts. *Tribol. Int.* **39**, 1719–1725 (2006)
- D.Y. Hua, L. Qiu, H.S. Cheng, Modeling of lubrication in micro contact. *Tribol. Lett.* **3**, 81–86 (1997)
- Y. Ju, T.N. Farris, Spectral analysis of two-dimensional contact problems. *ASME J. Tribol.* **118**, 320–328 (1996)
- S.C. Lee, N. Ren, Behaviour of elastic-plastic rough surface contacts as affected by the surface topography, load and materials. *STLE Tribol. Trans.* **39**, 67–74 (1996)
- S. Liu, Q. Wang, G. Liu, A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. *Wear* **243**, 101–111 (2000)
- N. Patir, H.S. Cheng, An average flow model for determining effects of three-dimensional roughness on partial hydrodynamic lubrication. *ASME Trans.* **100**, 12–17 (1978)
- N. Patir, H.S. Cheng, Application of the average flow model to the lubrication between rough sliding surfaces. *ASME Trans.* **101**, 220–229 (1979)
- C.H. Venner, EHL film thickness computations at low speeds: risk of artificial trends as a result of poor accuracy and implications for mixed lubrication modeling. *Proc. IMechE. Part J. J. Eng. Tribol.* **219**, 285–290 (2005)
- Y. Wang, Q.J. Wang, C. Lin, F. Shi, Development of a set of Stribeck curves for conformal contacts of rough surfaces. *STLE Tribol. Trans.* **49**, 526–535 (2006)
- D. Zhu, On some aspects of numerical solutions of thin-film and mixed elastohydrodynamic lubrication. *Proc. IMechE. Part J. J. Eng. Tribol.* **221**, 561–579 (2007)
- D. Zhu, Y.Z. Hu, The study of transition from elastohydrodynamic to mixed and boundary lubrication, in *Proceedings of STLE/ASME H.S. Cheng Tribology Surveillance*, pp. 150–156 (1999)

Mixed Metal Oxide Coatings

- Chemical Conversion Coatings

Mixture

- Lubricant Formulation

Modeling of Surface Roughness

- Surface Synthesis Based on Surface Statistics

Moderate Adhesion

- Gear EHL Film Thickness and Wear Risk Analysis

Modes of Fracture

ALAN T. ZEHNDER

Field of Theoretical and Applied Mechanics, Cornell University, Ithaca, NY, USA

Synonyms

Fracture symmetries; Tensile and shear cracks

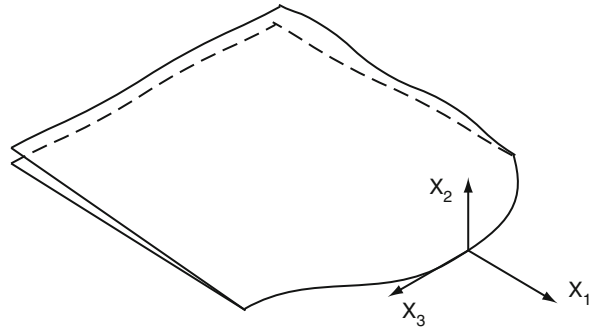
Definition

Modes of fracture refers to the decomposition of crack tip stresses into three loadings, or “modes.” The modes are Mode-I (stress orthogonal to the local plane of the crack surface), Mode-II (stress parallel to the crack surface but orthogonal to the crack front), and Mode-III (stress parallel to the crack surface and to the crack front).

Scientific Fundamentals

Consider the arbitrary fracture surface shown in Fig. 1. At any point on the crack front a local coordinate system can be drawn with the x_3 axis tangential to the crack front, the x_2 axis orthogonal to the crack surface, and x_1 orthogonal to the crack front. An observer who moves toward the crack tip along a path such that x_3 is constant will eventually be so close to the crack line front that the crack front appears to be straight and the crack surface flat. In such a case the three-dimensional fracture problem at this point reduces to a two-dimensional one. The effects of the external loading and of the geometry of the problem are felt only through the magnitude and geometry of the crack tip stress fields.

At the crack tip the stress field can be broken up into three components, called *Mode I*, *Mode II* and *Mode III* as sketched in Fig. 2 (Rice 1968). Mode I causes the crack to open orthogonal to the local fracture surface and results in



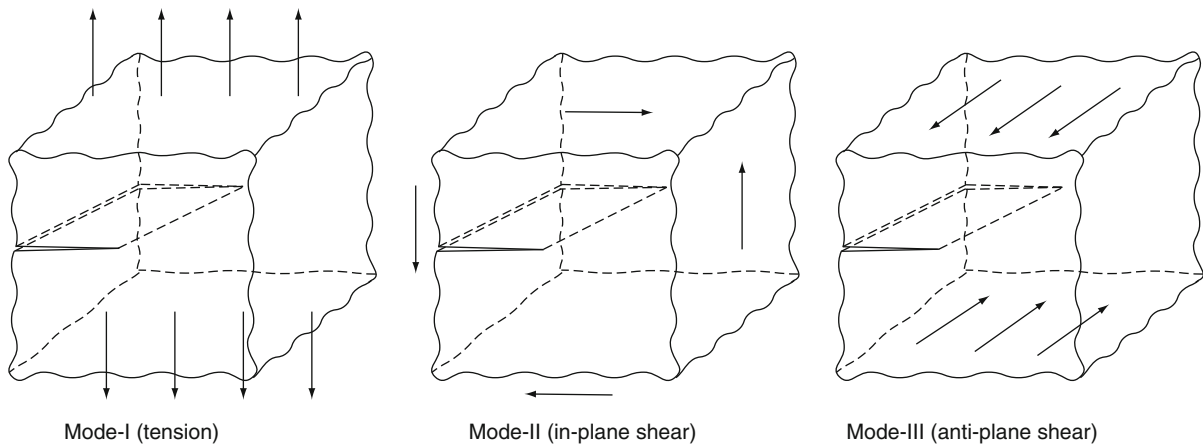
Modes of Fracture, Fig. 1 Crack front and crack surface for an arbitrarily shaped crack surface in a solid. At any point along the crack line a local coordinate system may be defined as shown

tensile or compressive stresses on surfaces that lie on the line $\theta = 0$ and that have normal vector $\mathbf{n} = \mathbf{e}_2$. Mode II causes the crack surfaces to slide relative to each other in the x_1 direction and results in shear stresses in the x_2 direction ahead of the crack. Mode-III causes the crack surfaces to slide relative to each other in the x_3 direction and results in shear stresses in the x_3 direction ahead of the crack.

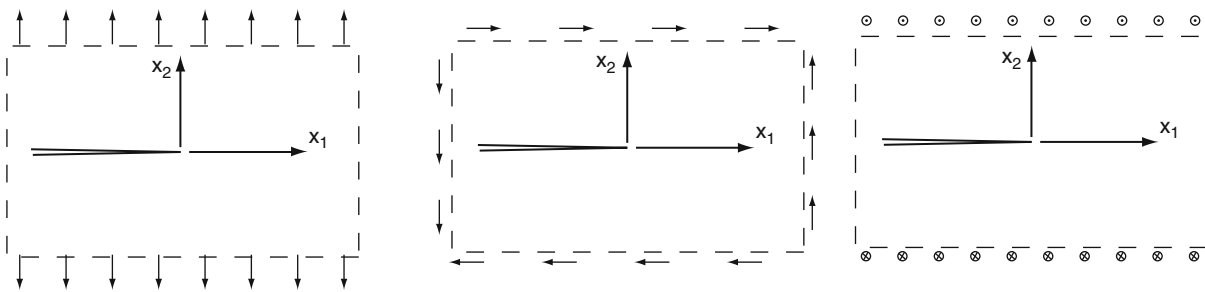
The solution of the crack tip fields can be broken down into three problems. Modes-I and -II are found by the solution of either a plane stress or plane strain problem and Mode-III by the solution of an anti-plane shear problem.

Another way to think of the three fracture modes is in terms of the symmetries of the displacement fields as shown in Fig. 3. In Mode-I the crack surfaces open up under tensile loading. In Mode-II the crack faces slide relative to each other in the x_2 direction. In Mode-III the crack faces slide relative to each other in the x_3 direction.

For linearly elastic materials the general crack tip stress fields are linear combinations of the fields corresponding to the three fracture modes. As an example of a problem in which all three modes of fracture are present, consider a circular crack subject to a far field loading stress of σ_∞ at an angle of β to the crack surface, as shown in Fig. 4. In this problem there is a constant Mode-I component of loading all along the crack front. However, the Mode-II and Mode-III components vary. On $\theta = 0$ and $\theta = \pi$ the crack front has Mode-I and Mode-II loading, but zero Mode-III loading. On $\theta = \pm\pi/2$, the crack front has Mode-I and Mode-III loading, but no Mode-II loading. In between, all three modes of loading are present.



Modes of Fracture, Fig. 2 Modes of fracture. Think of this as representing the state of stress for a cube of material surrounding part of a crack tip. The actual crack may have a mix of Mode-I, -II, and -III loadings and this mix may vary along the crack front. The equilibrating tractions on the front and back faces of a Mode-III cube are not shown



Modes of Fracture, Fig. 3 Symmetries of the crack tip displacement fields for each mode

Key Applications

The idea of modes of fracture allows for the simplification of the stress analysis, turning a complex three-dimensional problem into three two-dimensional problems whose solutions can be superimposed for general loadings. However, modes of fracture are also invaluable for the prediction of crack initiation, crack growth, and crack path evolution.

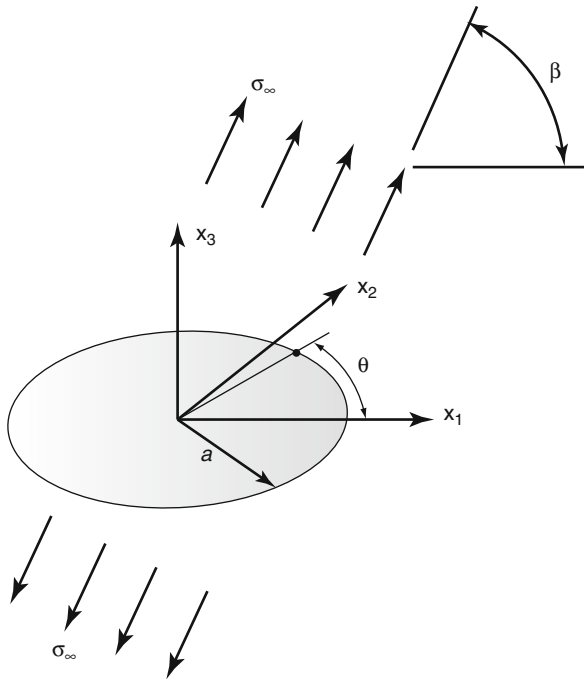
Due to the differences in the crack tip stress state for the three modes of fracture and due to possible contact and friction on the crack faces under the shear modes of fracture (Mode-II and Mode-III) the resistance of material to crack initiation and growth may be a function of the mode mixity, or relative values of Mode-I, -II, and -III loading.

In isotropic, homogeneous brittle materials, no matter what the loading, analysis and experiments show that cracks will generally evolve to a path such that the crack surfaces are perpendicular to the maximum principal

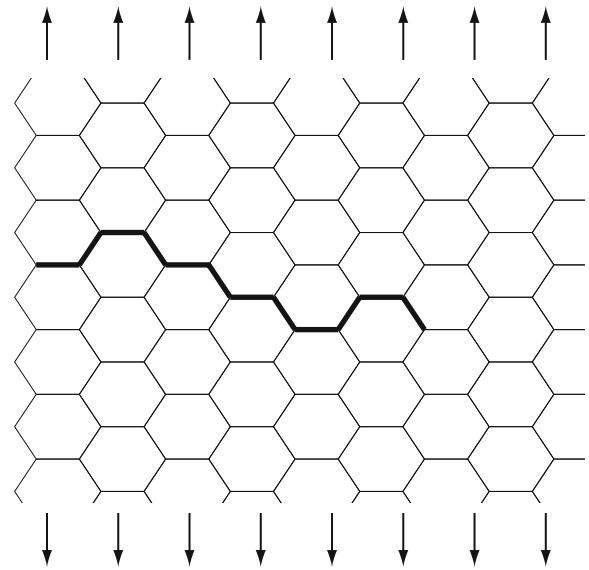
stress (if it is tensile) (Erdogan and Sih 1963). For example, in the simulation shown in Fig. 5 the crack is initially under pure Mode-II loading, but evolves along a path that puts the crack tip under pure Mode-I loading.

In polycrystalline materials, although the macroscopic crack path may evolve to be pure Mode-I, as the crack grows it may follow a path between grains (intergranular fracture) as sketched in Fig. 6. As it does so the local crack tip loading may include all three modes, the relative values of which will fluctuate as the crack follows its kinked path.

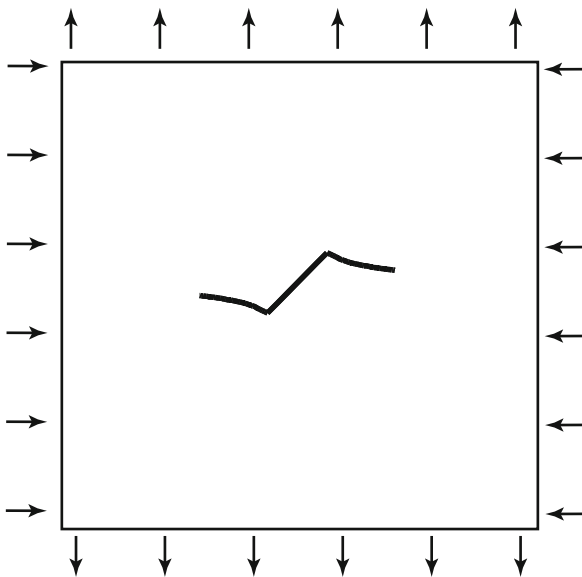
Cracks will generally grow under pure, or predominantly, Mode-II loading only if either the Mode-I loading is compressive or if the material contains weak planes that guide the growth of the crack regardless of the principal stress directions. In the case that the Mode-I load is compressive, suppressing tensile fracture, a crack growing under Mode-II loading will experience significant friction across the crack faces. This friction must be overcome for



Modes of Fracture, Fig. 4 Circular crack of radius a subject to uniform far field loading, σ_∞ at an angle of β to the crack surface. In the far-field, $\sigma_{33} = \sigma_\infty \sin^2 \beta$, $\sigma_{11} = \sigma_\infty \cos^2 \beta$, $\sigma_{13} = \sigma_\infty \sin \beta \cos \beta$. All other stress components are zero



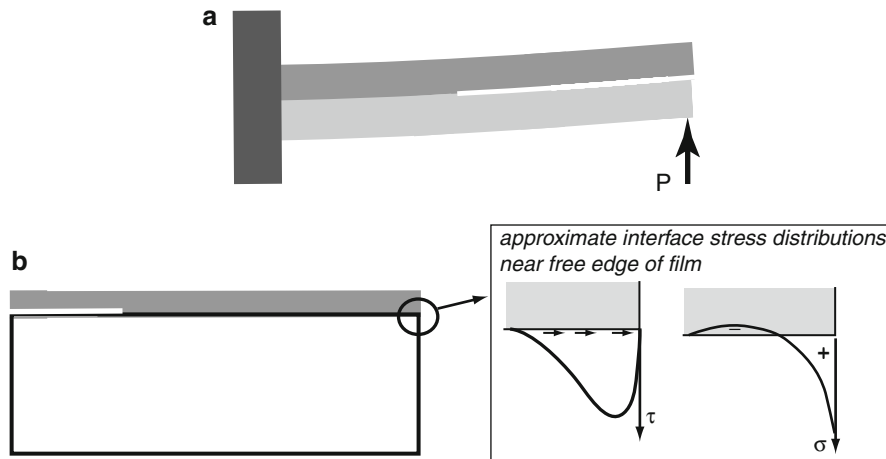
Modes of Fracture, Fig. 6 Schematic of intergranular crack growth in a polycrystalline material under Mode-I loading. As the crack grows, the mix of Mode-I, -II, and -III varies even if the macroscopic crack evolves to a pure Mode-I direction



Modes of Fracture, Fig. 5 Simulation of crack growth in a homogeneous, isotropic brittle material. Initial crack (straight line) is under pure Mode-II loading. The crack path evolves to be orthogonal to the maximum tensile principal stress so that the final crack tip is under pure Mode-I loading

the crack to propagate. Given the stick-slip nature of friction, crack growth under such conditions may consist of a number of dynamic fracture events (earthquakes are a familiar example).

Materials with weak planes abound. For example, composite structures are often built up of layers of fiber-reinforced polymer. The bond between the layers is generally much weaker than the strength of the layers in the fiber directions. Single crystal materials have different fracture resistances on their crystal symmetry planes. Biological materials such as wood and bone have strongly oriented microstructures that result in weak planes of fracture. In structures that are bonded or in systems containing a film or other thin layer bonded to a substrate, thermal mismatch stresses can cause a mix of Mode-I and -II fracture along the interface. For example, as shown in Fig. 7a, if a composite laminate containing a delamination due to impact damage or manufacturing flaws is subject to loading that causes transverse shear, the delamination crack will be under Mode-II loading. In a more general case of an arbitrarily shaped delamination the crack front will see a mix of Mode-II and -III loading. In the example of Fig. 7b, if a film bonded to a substrate has residual tensile or



Modes of Fracture, Fig. 7 Two examples of crack growth under Mode-II loading. **(a)** Composite laminate containing a delamination crack and subject to transverse shear loading. **(b)** A film on a substrate may crack under mixed-mode conditions along the interface of the film and substrate

compressive stresses, concentrated shear and normal stresses occur at the free edge of the film as shown schematically. These may be large enough to cause the film to debond from the substrate as illustrated on the left hand-side of the sketch. The debond crack tip will be under mixed mode loading (i.e., a combination of Mode-I and Mode-II).

Cross-References

► [Stress Intensity Factors](#)

References

- F. Erdogan, G.C. Sih, On the crack extension in plates under plane loading and transverse shear. *J. Basic Eng.* **85**, 516–527 (1963)
- J.R. Rice, Mathematical analysis in the mechanics of fracture, in *Fracture*, chapter 3, ed. by H. Liebowitz, vol. II (Academic Press, New York, 1968), pp. 191–311

Modes of Lubrication

► [Lubrication Regimes](#)

Modification of Elastomers

► [Polymeric Elastomers: Material Aspects of Tribology](#)

Modified UHMWPE for the Hip Joint (Particle Filled and Reinforced)

SHIRONG GE, QINGLIANG WANG

Institute of Tribology and Reliability, Department of Materials Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu, People's Republic of China

Synonyms

ALN – alendronate sodium; AW – glass-ceramic apatite-wollastonite; BHA – bovine bone hydroxyapatite; CF – carbon fiber; CNTs – carbon nano-tubes; E2 – 17 β -estradiol; HA – hydroxyapatite; HDPE – high-density polyethylene; NC – natural coral; PEEK – polyether-etherketone; UHMWPE – ultra-high molecular weight polyethylene

Definition

Ultra-high-molecular-weight polyethylene (UHMWPE or sometimes shortened to UHMW), also known as high-modulus polyethylene (HMPE) or high-performance polyethylene (HPPE), is a type of polyolefin made up of extremely long chains of polyethylene, which all align in the same direction, which can be classified as a linear homopolymer.

Polyethylene is a polymer formed from ethylene (C_2H_4). The generic chemical formula for polyethylene is $-(C_2H_4)_n-$, where n is the degree of polymerization. For an UHMWPE, the molecular chain can consist of as many as 200,000 ethylene repeat units. Put another way, the molecular chain of UHMWPE contains up to 400,000 carbon atoms (Kurtz 2004). The molecular weight of UHMWPE numbers is in the millions, usually between 2 and 6 million.

Structure

The non-static chains of UHMWPE imbue with internal (thermal) energy, and they can become mobile at elevated temperatures. When cooled below the melt temperature, the molecular chain of polyethylene has a tendency to rotate about the C-C bonds and create chain folds. This chain folding, in turn, enables the molecule to form local ordered, sheet-like regions known as crystalline lamellae. These lamellae are embedded within amorphous (disordered) regions and may communicate with surrounding lamellae by tie molecules, shown in Fig. 1. The lamellae are on the order of 10–50 nm in thickness, and 10–50 μm in length. The average spacing between lamellae is on the order of 50 nm (Kurtz 2004).

The weak bonding between olefin molecules allows local thermal excitations to disrupt the crystalline order of a given chain piece-by-piece, resulting in much poorer heat resistance than other high-strength polyethylene. Its melting point is around 144–152 °C (291–306 °F). It becomes brittle at temperatures below -150 °C (-240 °F). UHMWPE shows a white, opaque appearance at room temperature. At the temperature above the

melt temperature of the lamellae, UHMWPE becomes translucent.

The longer chain of UHMWPE serves to transfer load more effectively to the polymer backbone by strengthening intermolecular interactions, which results in the characteristic properties as following:

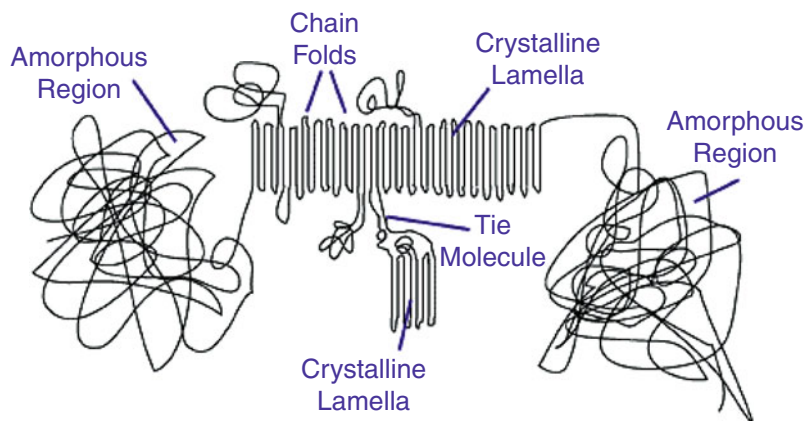
- Low coefficient of friction with self-lubricating performance
- High wear resistance
- Good chemical resistance
- Resistance to environmental stress cracking
- Dimensional stability over a wide temperature range
- High notched impact strength
- High energy absorption at high stress rates

Compression Molding

This manufacturing technique originated in Germany to manufacture medical grade UHMWPE product. The fundamentals apply to the compression molding of UHMWPE, where the consolidation of the materials is a direct result of the heat, pressure, and time combination used by the processors, this, however, is dependent upon the size of hydraulic presses/mold dimensions used.

Scientific Fundamentals

As one of the most potential methods to increase wear resistance of UHMWPE, the filling modification can effectively improve the hardness, deformation resistance, carrying capacity, and creep resistance of UHMWPE composites. Currently, UHMWPE composite filled with rigid particles and fibers have been widely reported in the literature.



Modified UHMWPE for the Hip Joint (Particle Filled and Reinforced), Fig. 1 Schematic of morphological features of UHMWPE chains

Ceramic or metal powders are the widely used reinforcing particles with higher hardness and wear resistance, such as Al_2O_3 , TiO_2 , SiO_2 , ZrO_2 , Ti, and Pt-Zr quasicrystal. These harder particles are highly dispersed in UHMWPE matrix, and then form the sub-microscopic structure of reinforcing particles enveloped by polymer. As the particles prevent the molecular chain movements caused by the matrix plastic deformation, the carrying capacity of the UHMWPE composite can be enhanced effectively. Therefore, addition of rigid particles to UHMWPE can produce a number of desirable effects; for example, an increase in stiffness, a reduction in coefficient of thermal expansion, and an improvement in creep resistance, fracture toughness, and wear resistance. It is evident that the improved of mechanical properties of UHMWPE composites are affected by a number of parameters, such as the size, shape, content, distribution of the reinforcing particles, and so on.

Another filling material is the rigid fibers, such as CF. The wear failure of these UHMWPE composites filled with fibers usually exhibit the fiber thinning due to wear, fracture of fibers, pulling-out, and peeling-off of fibers. Therefore, in order to assure the low wear rate of these composites, the fiber reinforcing materials should have the higher elastic modulus, interfacial bonding strength, and low coefficient of friction. The reasonable adjustment of fiber type, content, orientation, and spreading sequence can accommodate the requirements of low wear and high mechanical strength for UHMWPE composites. It's worth noting that the degree of orientation with respect to the applied stress is important for the fiber fillers.

Major factors affecting thermal, physical, and mechanical properties of UHMWPE composites reinforced by particles and fibers include the filler type and volume fraction, and interfacial state between the filler and UHMWPE matrix. Moreover, the other factors such as filler surface treatment and matrix-modification also play an important role (Wang and Bonfield 2001). In order to increase the interfacial strength between the two component phases and hence to improve mechanical properties of the resultant UHMWPE composite, silane coupling agents are normally used for the filler surface treatment to form the interfacial layer providing chemical adhesion for the fillers to matrix. Improved bonding strength achieved by both chemical adhesion and tight mechanical coupling can lead to improved mechanical properties of filling modified UHMWPE composites.

Particulate-Filled UHMWPE Composite

Two kinds of nano-scale and micron filling particles are commonly used to reinforce UHMWPE material.

Under the same conditions, the nano-particle can play a better strengthening effect due to a larger surface area. In general, the addition of micron particle about 10–20% (wt) is better, while only 5 % (wt) or less nano-particle added to UHMWPE can achieve the best enhancement. The reason is that the better bonding strength between the surface of nano-particle and UHMWPE matrix can take advantage of enhanced performance of the rigid particles.

Alumina nano-particle has higher hardness and elastic modulus, which can improved the hardness of UHMWPE/ Al_2O_3 composite significantly (Xiong et al. 2006). In the wear process, Al_2O_3 nano-particles dispersed in matrix of UHMWPE may produce the enrichment effect on the wear surface and act as the role of “lubricant.” Meanwhile, the enrichment alumina nano-particles were re-pressed into UHMWPE matrix under the positive pressure, which reduced the opportunity to the direct wear of UHMWPE and improved the wear resistance of UHMWPE/ Al_2O_3 composite. Carbon nano-tubes (CNTs) have the higher strength and mechanical properties. Only about 1.0 % (wt) of carbon nano-tubes added to UHMWPE can improve the modulus and impact strength of UHMWPE composites. But coefficient of friction has been reduced to some extent and wear resistance only had a little change compared with the pure UHMWPE. These results are closely related to the reinforcing efficiency of CNTs in composites, which depends strongly on the uniform dispersion of CNTs throughout the polymer matrix without destroying the integrity of the nano-tubes (Wang et al. 2005). In consequence, it remains to be seen how much of a wear reduction is possible with the use of filling particles in UHMWPE, and whether they degrade impact properties in the polymer.

UHMWPE/CF Composite

As a mainly reinforcing fiber, carbon fiber is constantly used to enhance the wear resistance of UHMWPE (Xiong 2005; Chen et al. 2010). Carbon fiber with high strength and elastic modulus can play the role of rigid load, and then prevent the embedding and plowing of wear particles. Not the wear resistance of UHMWPE/CF composites improves, but also the wear particles reduce obviously. Under dry friction and water lubrication, the main wear mechanisms on the worn surface of UHMWPE/CF composites are mainly characterized by the pulling-out and peeling-off of CF fibers. But, the typical wear features of wavy folds and plowing for the pure UHMWPE are not found. In addition, the chopped UHMWPE fibers have the same chemical characteristics as UHMWPE matrix, which can obtain the excellent mechanical properties of UHMWPE composite materials for their excellent

chemical compatibility and interfacial strength. Addition of short UHMWPE fibers can significantly increase the tensile strength, elastic modulus, impact strength and creep resistance of these self-reinforced UHMWPE composite, which is suitable for the load-bearing prosthesis (Chang et al. 2000; Jacobs et al. 2002). But, the UHMWPE composite reinforced by the oriented UHMWPE fiber is not ideal in bovine serum lubrication. In this UHMWPE material, layers of woven UHMWPE fibers are embedded in a UHMWPE matrix. It's obtained that this composite did not show any improvement on the wear rate compared to the standard UHMWPE due primarily to poor bonding between UHMWPE fibers and bulk matrix. Similar wear behaviors as the standard are found under the same conditions tested, with a wear rate of approximately $1 \times 10^{-7} \text{ mm}^3/\text{Nm}$. The average values of the coefficient of friction of all samples ranged from 0.08 to 0.11.

Therefore, further investigation is required to determine whether better matrix–fiber bonding will improve the wear performance of such self-reinforced composites.

UHMWPE/HA Composite

As a main inorganic component in human bones, hydroxyapatite is currently used as biomaterials for many applications in biomedicine because of its excellent biocompatibility and bioactivity. Up to now, some researches on UHMWPE/HA composites have also been reported (Fang et al. 2006, 2005; Wang et al. 2008). An increase in hydroxyapatite volume fraction leads to increases in strength, hardness and modulus of UHMWPE composites, with a simultaneous reduction in strain to failure. The tensile properties of UHMWPE/HA composites are compared with some well-studied bone-analogue composites and cortical bone as shown in Table 1 (Fang et al. 2006).

Modified UHMWPE for the Hip Joint (Particle Filled and Reinforced), Table 1 Mechanical properties of bone-analogue composites

Material	Processing method	Modulus (GPa)	Strength ^a (MPa)	Fracture strain (%)
HA/HDPE (20 vol%)	Twin-screw extrusion	1.60 ± 0.02^b	17.77 ± 0.09	34.0 ± 9.5
	Compression molding	1.81 ± 0.05^c	19.9 ± 0.07 (15)	39.7 ± 1.5
		1.55 ± 0.02^d	17.65 ± 0.17	80.0 ± 15.5
HA/HDPE (20 vol%)	Ultrasonic mixing	3.0 ± 0.1^e	25.7 ± 1.1 (24)	65.1 ± 9.9
	Compression molding	4.2 ± 0.1^f	29.6 ± 1.7	202.9 ± 66.6
HA/HDPE (50 wt%)	Twin-screw extrusion	4.0	39	35
	Injection molding			
HA/PEEK (20 vol%)	Shear mixing	~ 7.0	~ 70 (32)	<5
	Injection molding			
AW/HDPE (20 vol%)	Twin-screw extrusion	1.4 ± 0.1^g	(5.5 ± 0.9)	>20
	Compression molding	1.4 ± 0.1^h	(5.6 ± 0.5)	>20
HA/UHMWPE (20 vol%)	Twin-screw extrusion	6.8 ± 0.5	(26.6 ± 1.2)	375 ± 41
	Compression molding ball-milling	1.37 ± 0.06	11.5 ± 0.2^j	/
HA ⁱ /UHMWPE (30 wt%)	Compression molding ball-milled (swollen UHMWPE)	1.63 ± 0.07	12.0 ± 0.4^j	/
	Compression molding			
Cortical bone (50 vol%)	/	7–30	50–150	1–3

^aUltimate tensile strength and yield strength (marked in parentheses, measured from tensile curves in references)

^bWithout HA silanation and w/o PE grafting (HA: D0.5 = 4.14 μm)

^cWith HA silanation and w/PE grafting (HA: D0.5 = 4.14 μm)

^dWithout HA silanation and w/o PE grafting (HA: D0.5 = 7.32 mm)

^eSpherical HA

^fHA whiskers

^gAW size D0.5 = 4.5 mm

^hAW size D0.5 = 6.7 mm

ⁱWet synthesis and sintering at 900 °C for 2 h

^j $\sigma_{0.2}$

These composites exhibit higher Young's modulus than that of UHMWPE, while maintaining the excellent toughness feature of UHMWPE. The highest fracture strain of UHMWPE/HA composite reaches over 300%, significantly higher than other types of bio-composites. The particle size and morphology of hydroxyapatite have significant effects on the mechanical properties. The composites with smaller hydroxyapatite particles have a higher torsional modulus, tensile modulus and tensile strength. The resultant composite filled by nano-hydroxyapatite particles increases in Young's modulus by 90 % and in yield strength by 50 %. The addition of HA particles improve effectively the wear resistance of UHMWPE/HA composite. The largest reduction in wear rate is nearly 60 %, compared with those of unfilled UHMWPE. The abrasive and fatigue wear are the main wear mechanisms and the sizes of the wear particles become larger with an increase in addition of HA particles (Wang et al. 2008). In addition, bovine bone hydroxyapatite is a promising material for bone repairing and scaffold in bone tissue engineering. UHMWPE composites reinforced by bovine bone hydroxyapatite powders are prepared due to its better bioactivity, structure, and component similar to human bone (Wang et al. 2009; Knets et al. 1993). The addition of BHA particles to UHMWPE is beneficial for increasing the hardness and creep modulus of UHMWPE composites, which leads to the improvement of wear resistance. Under the bovine serum lubrication, the wear rate of UHMWPE/BHA composite cup decreased clearly by about 50 %, against CoCrMo femoral head in hip joint simulator under calf serum lubrication. HA (BHA) filling particles undergo plastic deformation under the applied load, and thus support a significant amount of the contact stresses during sliding. These behaviors reduce the effective shear stress on surrounding polymer and lead to the reduction in wear. As a result, preliminary results indicate a promising direction for the future development of HA/UHMWPE for biomedical applications.

UHMWPE/NC Composite

Natural coral possesses good biocompatibility because it has almost same composition as human bone. Products resulting from natural coral have been successfully applied in bone defect repairing. Natural coral implants in bone tissue is gradually absorbed and progressively replaced by newly formed bone. This behavior suggests that natural coral particles are friendly to human bone system and have potential for application in UHMWPE modification. Therefore, natural coral particles are added to UHMWPE to prepare the wear-resistant UHMWPE/NC composites (Ge et al. 2009). The experimental results show that the

addition of NC particles in UHMWPE improve the micro-hardness and scratch resistance of UHMWPE/NC composites, which increases in linear relation to natural coral contents. The wear rates of UHMWPE/NC composites decrease with the increasing contents of natural coral particles. The relation of wear mass loss of UHMWPE acetabular to the micro-hardness follows a negative power law. The wear mechanism of UHMWPE/NC composites is mainly controlled by adhesive wear. The investigation of UHMWPE wear debris reveals that the addition of NC particles in UHMWPE results in variations of size distribution of UHMWPE wear debris. Because micro-hardness of natural coral materials is much lower compared with CoCrMo alloy, few metal debris from CoCrMo ball are captured in the filter paper.

Crosslinked UHMWPE/ V_E Composite

The oxidative stabilization of crosslinked UHMWPE has become of primary relevance in the recent past. Currently the method of choice to improve the oxidation resistance of irradiated UHMWPE is melting, which decreases the concentration of residual free radicals to undetectable levels, but also decreases the crystallinity of polyethylene. The combination of reduced chain mobility and crystallinity decreases the resistance of irradiated UHMWPE to propagation of fatigue cracks (Oral et al. 2009). An alternative approach to improve the oxidation resistant of crosslinked UHMWPE is to add an antioxidant into irradiated UHMWPE, such as vitamin E. V_E can prevent the cascading oxidation reaction in UHMWPE by donation hydrogen to the primary alkyl free radicals on polyethylene chains and the peroxy free radicals caused by reactions with oxygen. Several aspects of such V_E -stabilized crosslinked UHMWPE are already published. Mechanical properties, fracture mechanisms and wear behavior in hip simulator have been investigated (Wang et al. 2010; Oral et al. 2008). V_E -stabilized crosslinked UHMWPE show the improved mechanical properties and wear resistance compared to crosslinked UHMWPE because the decrease in crystallinity caused by irradiating and remelting is avoided. The hypothesis is that additional increase in the crystallinity of UHMWPE will increase the strength and fatigue resistance of crosslinked UHMWPE. Therefore, it can be considered that V_E may be possible to improve the fatigue strength of crosslinked UHMWPE thereby obtaining a wear resistant and oxidation resistant joint bearing surface. This is promising for the use of alternate bearing surfaces made of UHMWPE in total hip and especially total knee arthroplasty, because fatigue damage and wear behavior under adverse loading conditions remain major concerns.

Drug-Loaded UHMWPE Composite

Recently, a novel drug-loaded UHMWPE composite has been reported by Qu et al. (Liu et al. 2009; Qu and Yang 2009). It takes full advantage of the principle of drugs blocking osteolysis and provides a new way to solve the problem of clinical loosening. It has been well reported that the UHMWPE wear particles are the major concern for the osteolysis, which finally lead to the aseptic loosening after total hip arthroplasty. General approach is to improve the wear resistance of UHMWPE acetabular cup, which has shown to be improved significantly by crosslinking, ion implantation, and filler modification. These researches have led to considerable improvements in reducing wear. However, the osteolysis remains an issue in clinical performance of hip prostheses. The aim of drug-loaded UHMWPE composites is expected that the drugs in UHMWPE composite acetabular cup could be released with the inevitable UHMWPE wear particles under the normal wear condition. Thus, the released drugs would directly or indirectly decrease the bone resorption of the osteoclasts, and prevent the artificial joint loosening finally. Due to the hot press temperature of UHMWPE to be at least at 180 °C, the drugs added into the UHMWPE should have a good thermostability. ALN and E2, with the potential application on osteolysis treatment and the high melting point, are added into UHMWPE powders to produce UHMWPE/ALN and UHMWPE/E2 composites through hot press. The present preliminary studies show that the processing of UHMWPE-ALN and UHMWPE-E2 composites with good mechanical and wear properties would be feasible. The heat processing of these composites would not alter the functional groups of ALN and E2. There are no significant differences in the density, hydrophobicity, crystallinity, and mechanical properties between the UHMWPE/ALN or UHMWPE/E2 and the standard UHMWPE. The most important is that the friction and wear performance would not be affected by adding ALN or E2 at the present dosage. The wear mechanisms of these composites are abrasive wear under dry friction, which is the same as the standard UHMWPE. As a result, whether UHMWPE/ALN or UHMWPE/E2 possesses the approving mechanical properties and wear performance compared with the standard UHMWPE, which might be used as the potential implanted drug carrier to prevent the particle-induced osteolysis in joint replacements.

Key Research Findings

Certain progress has been made in improving the wear resistance of UHMWPE from these researches. The most direct advantage of using the filling particles or fibers over

the crosslinking modification is the avoidance of free radical generation and accelerated oxidation. However, it remains to be seen some technical issues, such as the dispersion of particles (nano-scale), interfacial strength and precise control of morphology and size of particles or fibers. Especially, an immense challenge in using fillers in UHMWPE is the biocompatibility and bioactivity on the basis of considering the toxicological effects that the filler particles may have on the body. It's a pity that whether the biocompatibility or bioactivity of fillers has not been concerned universally. In fact, it should include two implications to improve the wear resistance of UHMWPE joint prostheses. The first is to enhance wear resistance to decrease the production of wear particles. The second is to increase the biocompatibility and bioactivity to decrease the adverse biological reactions of wear particles. However, most particulate or fiber fillers in UHMWPE for increasing the wear resistance are of inert materials. During the wear process, the spalling particles inevitably generate the toxicological effects on the body or cause the biological reactions with the biological tissues. Therefore, the new UHMWPE composites combined the biocompatible and bioactive filling reinforcements and drug modified UHMWPE matrix, can be obtained the double modification effects in improving wear resistance and reducing biological reactions of wear debris, which is an important direction of UHMWPE composites for future development.

Key Applications

As discussed in the previous paragraphs, the use of UHMWPE composites for hip replacement and augmentation provide appropriate stiffness and toughness and wear resistance, which is conducive to the improvement of the stability of the implants. The hard particle filled and reinforced gives stiffness and bioactivity to the composite, such as hydroxyapatite. Polyethylene is bioinert but tough and the composite has a stiffness similar to that of cortical bone. The material may be shaped or drilled and when implanted in the body produces a higher wear resistance than polyethylene alone. UHMWPE/HA composite has been in use as an orbital implant for either orbital floor fractures or volume augmentation (Kurtz et al. 1999). Vitamin E has been used to improve the oxidation resistance of cross-linking UHMWPE. Mechanical properties, fracture mechanisms, and wear behavior in hip simulator have been investigated in detail. But it is worth noting that the application of UHMWPE/CF composite, known as Poly IITM, is not satisfactory (Tanner et al. 1994). Expectations are thus that the Poly IITM would be more resistant to the pitting and delamination often seen in knee

arthroplasties. Wear testing conducted by the manufacturer reveals it to have an order of magnitude less wear than conventional UHMWPE, suggesting that the strength benefits would result in longer-lasting hip arthroplasties as well. However, pin-on-disk wear tests by McKellop show higher wear rates for Poly IITM than conventional UHMWPE, as well as abrasion of the metal by the exposed carbon fibers. Unfortunately, the promise shown by Poly IITM in the laboratory was not borne out in the clinical setting, and within a short time after implantation many patients presented with osteolysis and complete mechanical failure of their tibial bearing inserts.

Cross-References

► [Biotribology](#)

References

- N. Chang, A. Bellare, R.E. Cohen et al., Wear behavior of bulk oriented and fiber reinforced UHMWPE. *Wear* **241**, 109–117 (2000)
- R. Chen, Y.Z. Bin, Y. Nakano et al., Effect of chemical crosslinking on mechanical and electrical properties of ultrahigh-molecular-weight polyethylene-carbon fiber blends prepared by gelation/crystallization from solutions. *Colloid Polym. Sci.* **288**, 307–316 (2010)
- L.M. Fang, Y. Leng, P. Gao, Processing of hydroxyapatite reinforced ultra high molecular weight polyethylene for biomedical applications. *Biomaterials* **26**, 3471–3478 (2005)
- L.M. Fang, Y. Leng, P. Gao, Processing and mechanical properties of HA/UHMWPE nanocomposites. *Biomaterials* **27**, 3701–3707 (2006)
- S.R. Ge, S.B. Wang, X.L. Huang, Increasing the wear resistance of UHMWPE acetabular cups by adding natural biocompatible particles. *Wear* **267**, 770–776 (2009)
- O. Jacobs, M. Kazanci, D. Cohu et al., Creep and wear behavior of ethylene-butene copolymer reinforced by ultrahigh-molecular weight polyethylene fibers. *Wear* **253**, 618–625 (2002)
- I.V. Knets, L.O. Bunina, V.V. Filipenkov, Ultrahigh-molecular weight polyethylene and hydroxylapatite-based materials for replacement of bone tissue. *Mech. Compos. Mater.* **29**, 181–189 (1993)
- S.M. Kurtz, *The UHMWPE Handbook: Ultra-High Molecular Weight Polyethylene in Total Joint Replacement*, Elsevier Science & Technology Books (Academic, Boston, 2004)
- S.M. Kurtz, O.K. Muratoglu, M. Evans et al., Advances in the processing, sterilization, and crosslinking of ultra-high molecular weight polyethylene for total joint arthroplasty. *Biomaterials* **20**, 1659–1688 (1999)
- A.Q. Liu, S.X. Qu, M.M. Chao et al., UHMWPE carrying estradiol to treat the particle-induced osteolysis – processing and characterizing. *J. Biomed. Mater. Res. A* **90**, 496–505 (2009)
- E. Oral, C. Godleski Beckos, A.S. Malhi et al., The effects of high dose irradiation on cross-linking of vitamin E-blended UHMWPE. *Biomaterials* **29**, 3557–3560 (2008)
- E. Oral, A. Christine, C. Godleski Beckos et al., Improved resistance to wear and fatigue fracture in high pressure crystallized vitamin E-containing ultra-high molecular weight polyethylene. *Biomaterials* **30**, 1870–1880 (2009)
- S.X. Qu, D. Yang, Study on mechanical properties of UHMWPE carrying alendronate. *J. Med. Biomech.* **24**(5), 338–342 (2009)
- K.E. Tanner, R.N. Downes, W. Bonfield, Clinical applications of hydroxyapatite reinforced materials. *Br. Ceram. Trans.* **93**(3), 104–107 (1994)
- M. Wang, W. Bonfield, Chemically coupled hydroxyapatite-polyethylene composites: structure and properties. *Biomaterials* **22**, 1311–1320 (2001)
- Y.P. Wang, R.L. Cheng, L.L. Liang et al., Study on the preparation and characterization of ultra-high molecular weight polyethylene-carbon nanotubes composite fiber. *Compos. Sci. Technol.* **65**, 793–797 (2005)
- Q.L. Wang, D.K. Zhang, S.R. Ge, Biotribological behavior of ultra high molecular weight polyethylene composites containing coralline hydroxyapatite in a hip joint simulator. *Acta Materiae Compositae Sinica* **25**(2), 84–90 (2008) [In Chinese]
- Q.L. Wang, J.L. Liu, S.R. Ge, Study on biotribological behavior of the combined joint of CoCrMo and UHMWPE/BHA composite in a hip joint simulator. *J. Bionic Eng.* **6**(4), 378–386 (2009)
- S.B. Wang, Z.F. Ni, S.R. Ge, Tribological behavior of irritation cross-linked UHMWPE under dry sliding. *Adv. Mater. Res.* **97–101**, 605–609 (2010)
- D.S. Xiong, Friction and wear properties of UHMWPE composites reinforced with carbon fiber. *Mater. Lett.* **59**, 175–179 (2005)
- D.S. Xiong, J.M. Lin, D.L. Fan, Wear properties of nano-Al₂O₃/UHMWPE composites irradiated by gamma ray against a CoCrMo alloy. *Biomed. Mater.* **1**(3), 175–179 (2006)

Molecular Dynamics Simulation of Hydrophobic Interaction and Hysteresis at Different Length Scales

M

YAJIE LEI, YONGSHENG LENG

Department of Mechanical & Aerospace Engineering,
The George Washington University, Washington,
DC, USA

Synonyms

PMF – potential of mean force

Definitions

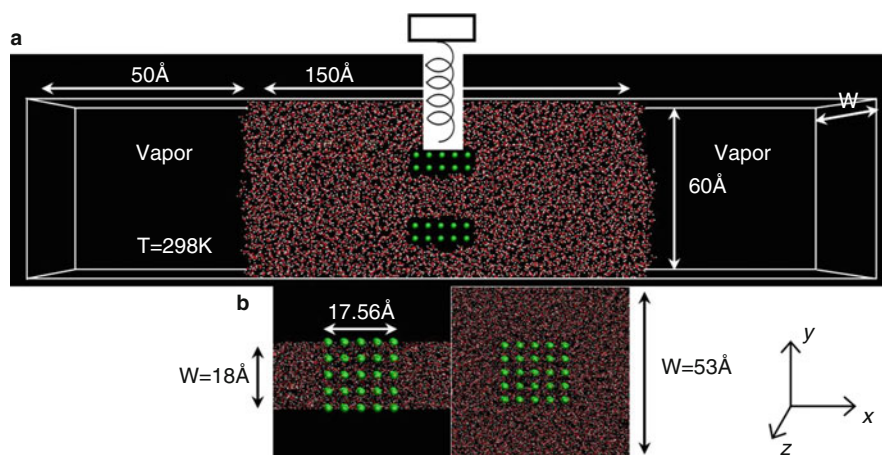
Hydrophobic interaction is the tendency of hydrocarbon-like groups in solutes to form intermolecular aggregates in an aqueous medium. Hydrophobic attractive forces are usually associated with this process.

Hydrophobic hysteresis is a force or free energy hysteresis during the separation between two hydrophobes in aqueous solution.

Scientific Fundamentals

Mechanism

Hydrophobic interaction between apolar species in water is a fundamental question in surface and interfacial



Molecular Dynamics Simulation of Hydrophobic Interaction and Hysteresis at Different Length Scales, Fig. 1 Illustration of the simulation model

science. Despite the basic principles underlying the hydrophobic effect being qualitatively well understood, only recently have theoretical developments begun to explain and quantify many features of this ubiquitous phenomenon (Chandler 2005). For example, theory predicts that the hydrophobicity (the spontaneous drying and exclusion of water molecules) at hydrophobe-water interfaces depends on the length scale of apolar species (Lum et al. 1999).

Hydrophobic attraction can be measured through the surface force experiment. This was accomplished as early as the 1980s by Israelachvili and Pashley (1982). To understand the molecular mechanism of hydrophobic interaction, molecular dynamics (MD) simulation is performed to mimic the force measurement procedure through a driving spring model (Lei and Leng 2012). Two rectangular thin hydrophobic plates are immersed in a large water bath (Fig. 1a). The normal approach and retraction of the upper hydrophobic plate relative to the lower one is driven by a spring in the vertical direction. As the separation distance between the two plates is gradually decreased, spontaneous hydrophobic drying can be observed. The hydrophobic attractive forces between the two plates can be directly obtained from the elongation of the driving spring tethered between the upper plate and the driving block. The water bath has two liquid–vapor (LV) external interfaces along the horizontal direction. These two LV interfaces are well maintained during the MD simulation, keeping the pressure of the system close to the ambient condition.

Example

The simulation model is applied to two systems in order to understand the size dependence of hydrophobic drying and hysteresis. Model I contains two penetrated plates along the z -direction, equivalent to infinitely long parallel plates along the z -direction (see the left panel in Fig. 1b). Model II contains two finite rectangular plates parallel to each other (see the right panel in Fig. 1b). The simulation box lengths along the z -direction are 18 Å and 53 Å for the two cases, while the dimensions in the y - and x -directions are 60 Å and 250 Å, respectively. System I has 4,574 TIP4P water molecules (Jorgensen et al. 1983), while system II has 12,864 water molecules. Each hydrophobic plate is composed of $5 \times 5 \times 2$ nonpolar monomers that represent CH_3 groups. Note that systems I and II have the same shortest characteristic dimension of 17.56 Å in the x -direction. The area per monomer on the hydrophobic surface is around 0.19 nm^2 , corresponding to the area of one CH_3 tail group on the hydrophobic paraffin surface. The OPLS (optimized potentials for liquid simulation) protein model for CH_3 group is used for all monomers in the hydrophobic plates. The Lennard-Jones parameters are given by $\epsilon_{\text{CH}_3} = 0.16 \text{ Kcal/mol}$ and $\sigma_{\text{CH}_3} = 3.91 \text{ Å}$. The interaction between hydrophobic monomers and the TIP4P water is described by the Lorentz-Berthelot combining rule (Allen and Tildesley 1987). In the MD simulations, short-range interactions are truncated at 9 Å and electrostatic interactions are calculated by the particle-particle particle-mesh (PPPM) method (Luty et al. 1994). The distance between CH_3 groups in hydrophobes

is 4.39 Å, which is equal to $2^{1/6}\sigma_{\text{CH}_3}$. Periodic boundary conditions (PBC) are applied in the three directions. The temperature is controlled at 298 K by the Nosé-Hoover thermostat (Allen and Tildesley 1987).

State-of-the-Art Findings

Hydrophobic Attractive Force and Hysteresis

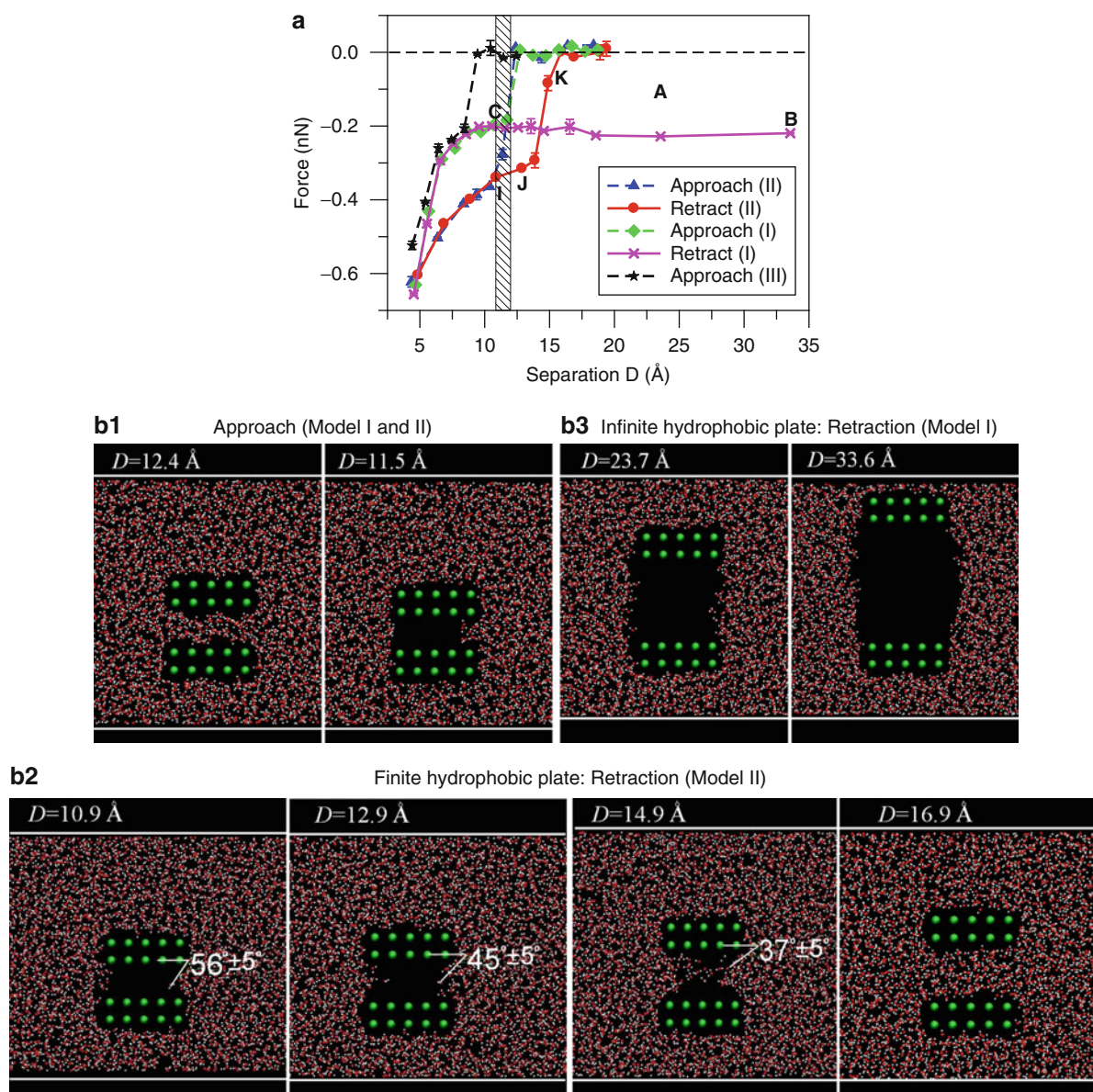
Starting from the equilibrium separation $D = 19.4$ Å, as shown in Fig. 2a, the normal approach and retraction processes between the two hydrophobic plates are simulated by pushing or pulling the driving spring. For every small driving distance (usually within 1–2 Å near the hydrophobic drying transition state), the hydrophobic attractive force is calculated by averaging the spring force over the 8 ns simulation time. Fig. 2a shows the attractive forces versus the surface separation D during the normal approach and retraction for the two systems I and II. At larger surface separations, the force slightly fluctuates around zero. When the separation is below a critical distance around $D_c = 11.5$ Å, the attractive forces rapidly increase for both systems, accompanied by spontaneous hydrophobic drying. At this critical distance, the infinite system (model I) equilibrates at a relatively smaller attractive force (~ 0.18 nN) compared with the larger attractive force (~ 0.28 nN) observed for the finite system (model II). This force magnitude difference at the initial drying process is related to the difference of the system geometry. Following the initial hydrophobic drying, the attractive hydrophobic forces for both systems continues to increase until hydrophobic collapse, while the force magnitude of the finite plate (model II) increases much faster than that of the infinite plate (model I). The hydrophobic force-distance curves for the two systems provide very detailed information on the force law from spontaneous drying to hydrophobic collapse. Associated with the sudden increase of the attractive forces, the density of the confined water fluctuates dramatically. The corresponding transition regime is marked as a shaded area in Fig. 2a between 11 and 12 Å. The snapshots of the nanobubble formation during the transition (point C in Fig. 2a) are shown in Fig. 2b1 for both systems. Since the two systems have the same shortest characteristic length along the x -direction (i.e., 17.56 Å), the result demonstrates that the critical distance for the spontaneous drying depends on the shortest characteristic dimension of the hydrophobic surface. To further test this idea, an additional MD simulation is performed for the hydrophobic interaction between two $3 \times 5 \times 2$ hydrophobes. In this case the characteristic dimension along the x -direction is 8.78 Å, but the dimension along the z -direction is infinitely long (this system is

called Model III). The spontaneous drying occurs at an even shorter distance (~ 9.5 Å). The variation of the hydrophobic attractive force is also shown in Fig. 2a for comparison.

Following the rapid increase in attractive force and spontaneous drying, the two surfaces jump into contact at $D = 4.39$ Å, which is the nearest-neighbor distance between CH_3 groups in the hydrophobes. From this contact state, hydrophobic hysteresis is simulated by pulling the driving block backward. The force-distance curves during retraction are shown in Fig. 2a. It is seen that when D is less than D_c no hysteresis occurs. When D is larger than D_c , significant force hysteresis arises. For the finite hydrophobic plate (Model II), the hysteresis occurs starting at point I and ending at point K. At these different stages, the geometric angle between the liquid–vapor and solid–vapor interfaces within the nanobubble is shown in Fig. 2b2. For the infinite hydrophobic plate (Model I), Fig. 2a shows that a much longer force hysteresis is observed that levels off at -0.22 nN towards $D = 33.6$ Å. This corresponds to a much larger internal vapor phase between the two infinite hydrophobic plates. The snapshots of the nanobubble at points A and B in Fig. 2a are shown in Fig. 2b3. This nanobubble can sustain over many nanoseconds MD relaxation without collapse, and it is so stable that, due to the yet small simulation system, it is difficult to see at what distance it will collapse. This is in contrast to the finite hydrophobic system (Model II) in which a nanobubble with a maximum height of 15 Å can sustain for a period of time. The above results indicate that the hydrophobic hysteresis depends on the longest characteristic dimension of the hydrophobe (here for model I, the dimension along the z -direction is infinite).

Variations of the Potential of Mean Force

The mechanism of hydrophobic drying and hysteresis can be explored by further calculating the potential of mean force (PMF) along the paths of approach and retraction. This quantity is a measure of the free energy change that determines the strength and likelihood of the association between two hydrophobic plates in water. The PMF is evaluated by means of the integration of the attractive force over the displacement of the external driving spring, projected onto the normal approach and retraction direction. The results are obtained by averaging over four independent trajectories along the same pulling path. Figure 3 shows the PMF per unit surface area as a function of separation D for the finite and infinite hydrophobic plates (Model II and I). The zeros of PMF are taken at the minimum equilibrium separation $D = 4.5$ Å. For the Model II system, as the plates are brought closer, from



Molecular Dynamics Simulation of Hydrophobic Interaction and Hysteresis at Different Length Scales, Fig. 2 (a) Attractive hydrophobic forces as a function of surface separation D for the $5 \times 5 \times 2$ infinite plate (Model I) and finite plate (Model II) during approach and retraction. The shaded area around the critical separation distance $D_c = 11.5$ Å represents the transition regime of drying and nanobubble formation for the two systems (green and blue). The same transition regime for the two systems indicates that the unstable critical distance depends on the shortest characteristic dimension of the plate. This is further supported by the $3 \times 5 \times 2$ infinite hydrophobes interaction (Model III) in which hydrophobic drying occurs at $D_c = 9.5$ Å. Error bars in the figure represent the force variances. The standard deviation is calculated by block averaging over a total of eight 500 ps windows. (b) Snapshots of nanobubble during approach-retraction process between two hydrophobic plates. All configurations show the equilibrium states after 8 ns relaxation. Here (b1) shows hydrophobic drying transition during the normal approach in Model I and II systems; (b2) shows the retraction snapshots in Model II immediately after the hydrophobic collapse. The variation of the geometric angle between the liquid-vapor and solid-vapor interfaces within the nanobubble corresponds to the points I, J, and K in Fig. 2a (b3) shows the retraction snapshots in Model I, corresponding to points A and B in Fig. 2a

LVMD – liquid–vapor molecular dynamics

Definitions

Simple nanoconfined fluids refer to a type of nonpolar spherical molecular liquids under nanometers confinement, such as octamethylcyclotetrasiloxane (OMCTS) confined between two molecularly smooth mica surfaces, or a typical model system of argon liquid confined between two crystalline solids.

Liquid-to-solid phase transition is a physical phenomenon in which a simple nanoconfined fluid forms a layered, solid-like structure that can support a finite shear stress.

Solvation force is a specific type of force between two solids immersed in liquid. In nonpolar fluids, this type of force exhibits an oscillatory behavior due to the recrystallization-melting layering transition of the simple nanoconfined fluids.

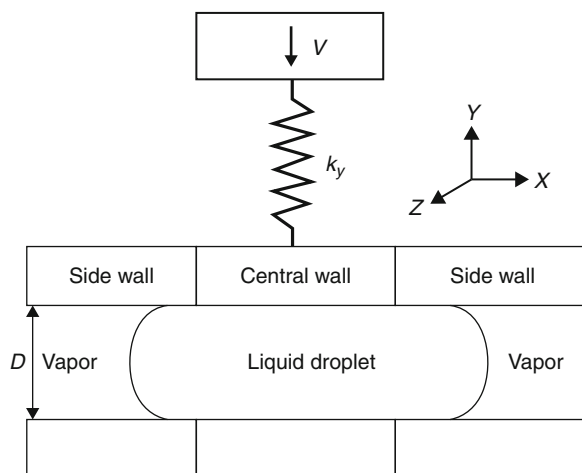
LVMD is a molecular dynamics method in which liquid-vapor interfaces are maintained in molecular ensemble.

Scientific Fundamentals

Mechanism

In surface force experiments, simple nonpolar fluids such as octamethylcyclotetra-siloxane (OMCTS) confined between two molecularly smooth mica surfaces may undergo a liquid-to-solid phase transition when confined to a film thickness around 6–7 molecular diameters [Klein and Kumacheva 1995, 1998]. It turns out that this is a very controversial issue in the surface force experimental community, since other experiments found that OMCTS underwent either glass transition (Demirel and Granick 1996) or extensive force oscillations (Zhu and Granick 2003), depending on the mica preparation.

This fundamental question has been explored recently by a LVMD simulation technique. The main features of LVMD [Leng 2008; Lei and Leng 2010] are shown in Fig. 1. A simple driving spring model is applied to mimic the force measurement in surface force experiments. The coexistence of two vapor phases around the central liquid droplet allows for the squeeze-out of the fluid to proceed. The lateral pressure is maintained at vanishingly low values due to the existence of vapor phases. In LVMD simulations, the normal approach and retraction of the upper wall is controlled by a driven dynamics of the driving system (Fig. 1). The hydrodynamic effect is eliminated when the rate of normal compression is sufficiently low. Only the forces acting on the upper central wall are counted towards the total normal force. In this way, when the liquid droplet is squeezed out sufficiently,



Molecular Dynamics Simulation of Phase Transition and Solvation Force Oscillation of Simple Nanoconfined Fluids, Fig. 1 Illustration of the LVMD simulation

meniscus effect will have minimal effect on the forces between two central walls.

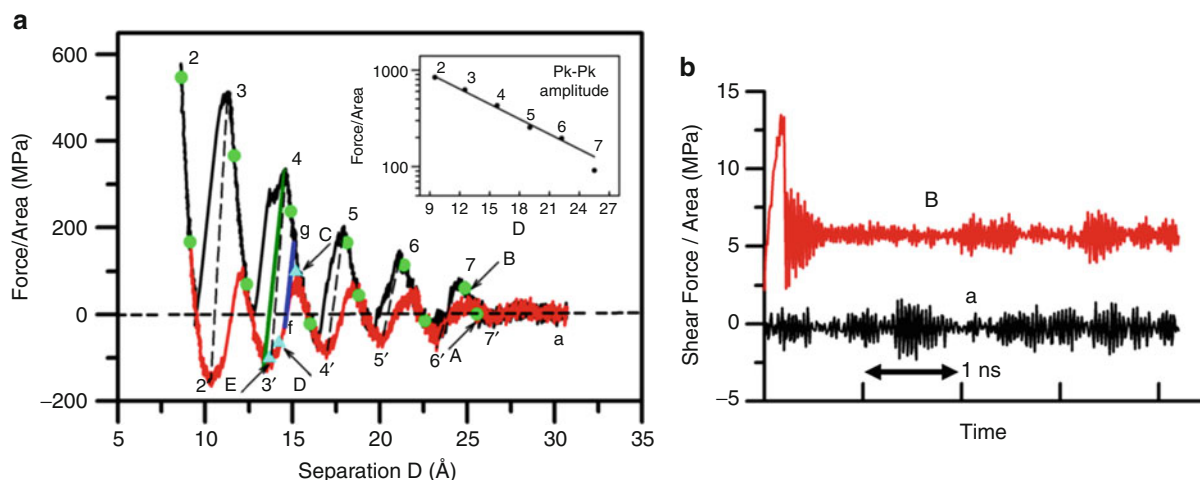
Example

Consider a simulation system consisting of a liquid droplet of argon and two face-centered cubic (f.c.c.) crystal walls with (111) surfaces. For liquid argon, force parameters in simple Lennard-Jones (LJ) atomic potential are $\epsilon = 0.2381$ kcal/mol and $\sigma = 0.3405$ nm. Each confining wall is composed of a central wall that has the same interaction strength as the argon-argon interaction (i.e., $\epsilon_{wf} = \epsilon_{ff} = \epsilon$, where w and f stand for wall and fluid, respectively), and two side walls on which the wall-fluid interaction is decreased to $\frac{1}{4} \epsilon$ to keep argon molecules in this region in a liquid phase. Assume $\sigma_{wf} = \sigma$ and the f.c.c. wall lattice constant $a = 0.5405$ nm, giving the first neighbor distance of f.c.c. wall particle as $a/(2)^{1/2} = 0.3822$ nm. A standard cutoff distance of 2.5σ is used in simulations. The sizes of MD simulation box along the x-, y-, and z-directions are 76.59, 11.42, and 3.71 nm, respectively. The central wall of interest has a length of 10.7 nm along the x-direction. The simulation system contains 4,319 liquid argon atoms and 8,823 solid wall atoms. The temperature of the system is controlled at 85 K, corresponding to a liquid state of argon.

State-of-the-Art Findings

Solvation Force Profile

For the argon system described above, the normal approach and retraction force-distance profiles between



Molecular Dynamics Simulation of Phase Transition and Solvation Force Oscillation of Simple Nanoconfined Fluids, Fig. 2

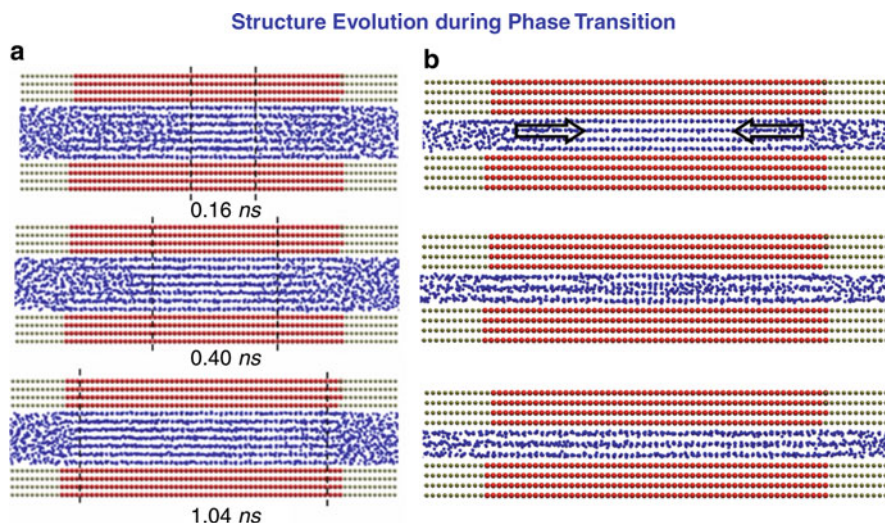
(a) Force-distance profiles between two solid surfaces in argon during normal approach (*black solid line*) and retraction (*red solid line*). The numbers n and n' in the figure correspond to the force maxima and minima at different layers. The *solid circles* A, B, etc. and *solid triangles* E, D, and C represent static forces during normal approach and retraction, respectively. The *dashed lines* show the unstable $n \rightarrow n - 1$ transition regions during normal approach, in which the force gradient $|\partial F / \partial D|$ is greater than the driving spring force constant. The lines 4-3' and f-g correspond to force relaxations of unstable transitions of $4 \rightarrow 3$ and $3 \rightarrow 4$, respectively. The inset shows the peak-to-peak amplitudes of the force oscillations as a function of D . The exponential decay length is estimated around 0.8 nm. (b) Variations of static shear forces across the seven-layer solidified film (point B in Fig. 2a) and a liquid film (point a in Fig. 2a) after an initial pulling of driving spring by 1 Å

two solid surfaces at a normal driving velocity $v = 0.05$ m/s is shown in Fig. 2. Forces are normalized by the contact area of the central wall. As can be seen in the figure, significant repulsive solvation force starts at seven layers ($n_c = 7$), corresponding to a film thickness of 2.5 nm or about seven molecular diameters of argon. This is attributed to an abrupt layering transition from a liquid phase to a solidified structure. Figure 2b clearly shows that the layered structure can sustain a finite shear stress (~ 6 MPa) after an initial pulling of 1 Å (force signal B in Fig. 2b), while the liquid phase cannot support any static shear stress with the same initial pulling (force signal a in Fig. 2b). Indeed, for subsequent $n = 6-2$ layered structures, further compression before $n \rightarrow n - 1$ layering transition always finds more ordered solid phase under higher pressures. The adhesion forces are obtained through retraction simulations, as represented by the number n' in Fig. 2a. The peak-to-peak amplitudes of force oscillations ($n \rightarrow n'$) decay dramatically as the separation D increases. The exponential decay length is estimated around 0.8 nm (see the inset in Fig. 2a). The unstable $n \rightarrow n - 1$ layering transitions during normal approach, in which the force gradient $|\partial F / \partial D|$ is greater than the driving spring force constant, are shown by the dashed lines in Fig. 2a. There are also unstable $n - 1 \rightarrow n$ transitions during

retraction, such as $3 \rightarrow 4$ shown in Fig. 2a. However, unlike the $n \rightarrow n - 1$ transition, in which the relaxed solvation forces can touchdown the adhesion valleys, forces during the $n - 1 \rightarrow n$ transition only reach a portion of the maximum repulsive peaks, such as $f \rightarrow g$ shown in Fig. 2a. Static forces exist even in the $3 \rightarrow 4$ unstable climbing regions during retraction. It is found that these forces correspond to an ordered three-layer structure (point E), a slightly less-ordered structure (point D), and an ordered four-layer structure (point C), respectively.

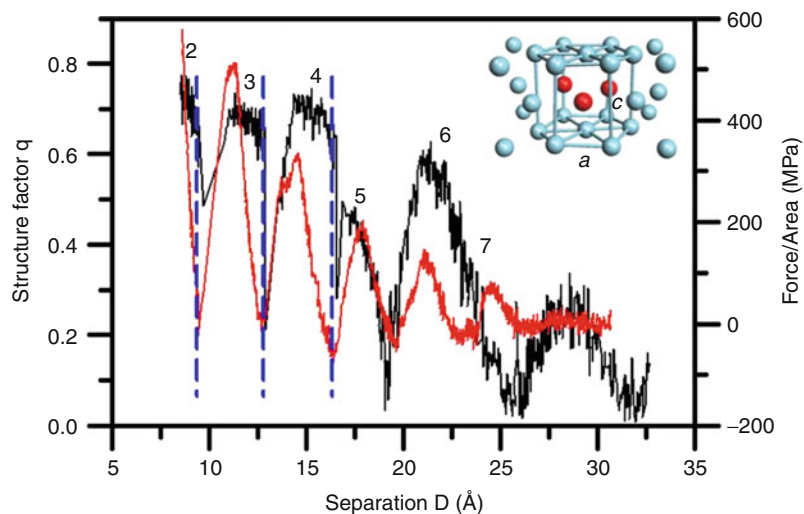
Structure Evolution During Phase Transition

A detailed MD animation analysis shows that the transition from a liquid phase to a layered solid phase at seven molecular diameters thickness is abrupt. The three panels in Fig. 3a show that at this critical distance, the nucleation of layered structure, starts at the central region. The layered structure grows and propagates outward to the edges of the central wall within 1.04 ns. A finite shear stress (Fig. 2b) can be sustained, demonstrating that the layered structure is indeed in a solid phase. When the solid phase of confined film is formed, further compression of the film before $n \rightarrow n - 1$ layer transition results in more compact ordered phase that can support an even higher pressure.



Molecular Dynamics Simulation of Phase Transition and Solvation Force Oscillation of Simple Nanoconfined Fluids, Fig. 3

(a) Liquid-to-solid phase transition at seven molecular layers during normal approach. At a critical distance of 2.6 nm, nucleation of solid phase starts at the central region. This solid phase grows and propagates outward to the edges of the central wall within 1.04 ns. (b) The unstable solid-liquid-solid transition ($4 \rightarrow 3$) during continuing approach. Arrows in the top panel indicate the inward shrinking of the solid phase



Molecular Dynamics Simulation of Phase Transition and Solvation Force Oscillation of Simple Nanoconfined Fluids, Fig. 4

Quantitative measure of the structure property of the solidified film. The figure shows the variation of the structure factor q of the confined film versus the wall separation during normal approach. The force profile is also shown in the figure. The force minima may correspond to very large q values (the blue dashed lines for $n = 2-4$), indicating that the solid phase is not induced by high pressure

The sudden pressure drop to a negative value corresponds to a complete melting of the solid phase, followed by a recrystallization of $n - 1$ layer. Figure 3b shows the $4 \rightarrow 3$ layer transition, which begins with a highly ordered

four-layer solid phase. An inward shrinking of solid phase (the top panel in Fig. 3b) takes place with an estimated speed around a few m/s. This inward squeeze-out, initiated from the solid-liquid phase boundary, is primarily

due to the maximum compressive stress at the edge of solid phase. Following the melting process completion, a recrystallization of a three-layer solid phase is accomplished (see the intermediate and lower panels in Fig. 3b).

Structure and Mechanical Properties of the Solidified Film

The ordered solidified phases adopt a hexagonal close-packed (h.c.p.) crystalline structure (see the inset in Fig. 4). The ratio of the lattice constants, c/a , varies from 1.59 to 1.75. For $n = 4$ layer, the lattice constants $a = 0.3758$ nm and $c = 0.6038$ nm. The c/a ratio is equal to 1.607, very close to the standard value of 1.633 for h.c.p. crystals. Figure 4 shows the variations of the structure factor q of the confined film versus distance during normal approach. The force maxima usually produce larger values of q due to the more compact and ordered phase in the crystallized film. However, the force minima also correspond to very large q . The key point is that the abrupt formation of the ordered solid phase is not pressure-induced. In contrast, this solid phase is purely confinement-induced.

The compressive elastic moduli during compression from 7 to 3 monolayers are estimated as 1.63, 2.45, 3.39, 3.74, and 5.13 GPa. The tensile elastic moduli during retraction from 3 to 6 monolayers are estimated as 1.41, 1.65, 1.31, and 1.46 GPa. Moreover, Fig. 2a shows that the peak amplitudes of repulsive oscillatory forces are always larger than those of attractive oscillatory forces. These indicate that the mechanical properties of the solidified h.c.p. film are highly asymmetric.

Key Applications

Surface and interfacial science; friction and lubrication; clay swelling; complex dynamics in biological systems; nanotribology.

Cross-References

- [Lubrication of Miniature Systems](#)
- [Surface Force Apparatus](#)
- [Surface Forces, Surface Tension, and Adhesion](#)
- [Thin Film Lubrication](#)

References

- A.L. Demirel, S. Granick, Glasslike transition of a confined simple fluid. *Phys. Rev. Lett.* **77**, 2261 (1996)
- J. Klein, E. Kumacheva, Confinement-induced phase-transitions in simple liquids. *Science* **269**, 816 (1995)
- J. Klein, E. Kumacheva, Simple liquids confined to molecularly thin layers. I. Confinement-induced liquid-to-solid phase transitions. *J. Chem. Phys.* **108**, 6996 (1998)

Y.J. Lei, Y.S. Leng, Force oscillation and phase transition of simple fluids under confinement. *Phys. Rev. E* **82**, 040501 (2010)

Y.S. Leng, Hydration force and dynamic squeeze-out of hydration water under subnanometer confinement. *J. Phys. Condens. Matter* **20**, 354017 (2008)

Y. Zhu, S. Granick, Reassessment of solidification in fluids confined between mica sheets. *Langmuir* **19**, 8148 (2003)

Molecular Gas Film Lubrication

SHENG SHEN, GANG CHEN

Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Definition

Molecular gas film lubrication is gas film lubrication under ultrathin clearance conditions, where the clearance is comparable to or smaller than the molecular mean free path of gas molecules. In molecular gas film lubrication problems, the gas cannot be treated as a continuum, and molecular rarefaction effects become important. Solving such lubrication problems requires using the kinetic theory of gases or rarefied gas dynamics.

Scientific Fundamentals

Gas film lubrication uses pressurized gas to provide non-contact support for bearing components with small clearances. One of the most distinctive properties of gaseous lubricants is their very low viscosity, which makes gas lubricated bearings virtually silent, low friction, vibration-free, and capable of working at extremely high rotational speeds. According to the method employed in generating the load-carrying pressurized gas film between the bearing surfaces, gas lubricated bearings can be classified as three types: self-acting, externally pressurized, and squeeze film (Powell 1970). Self-acting gas bearings generate their load capacity from the rotation of one of the bearing surfaces. The viscous shear-driven mechanism is similar to that of hydrodynamic oil bearings. Externally pressurized bearings are supplied with high-pressure gas from an external source, for example, a compressor. In squeeze film bearings, one of the bearing surfaces vibrates in the direction normal to the plane of bearing surfaces, thus, the pumping effect of the gas film generates pressure to support a load. The pressure distribution generated in gas bearings is calculated using the well-known Reynolds equation based on mass conservation (Hamrock et al. 2004). In the continuum limit, the mass flow rates are obtained from solving the

Navier–Stokes equations. Often, the fluid flow problem for gas film lubrication can be decomposed into the superposition of Poiseuille flow and Couette flow.

With advances in microfabrication technology, the gas film lubrication in many microelectro-mechanical systems (MEMS) devices works under ultrathin clearance conditions. For instance, in current hard disk drive designs, the clearance between the flying head slider and the magnetic disk is reduced to below 5 nm in an effort to achieve higher storage density. Such a small gap size is much smaller than molecular mean free path of air molecules, which is around 70 nm at one atmosphere pressure. Therefore, in gas film lubrication problems with ultrathin clearances, generally, the gas lubricants cannot be treated as a continuum, and hence molecular rarefaction effects become important (Bhushan 1995).

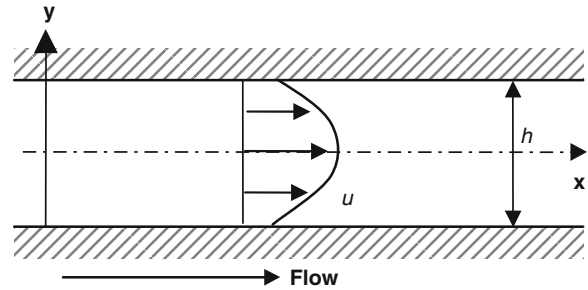
One most important parameters to evaluate the rarefaction of gas flows is the Knudsen number, Kn , which is defined as

$$Kn = \frac{\lambda}{h},$$

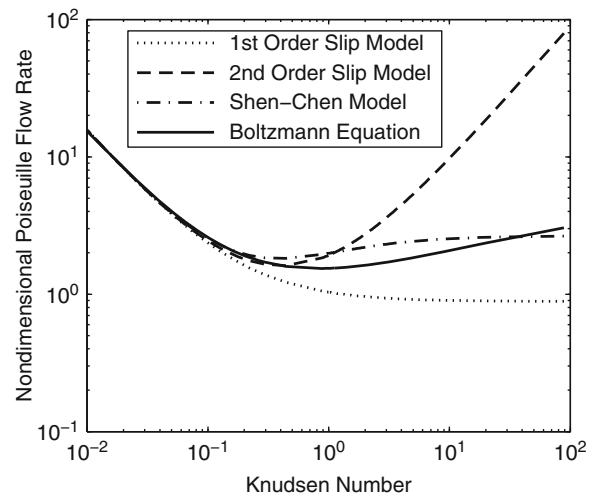
where λ is the molecular mean free path of gas molecules and h is the characteristic length of the flow (Kennard 1938). In terms of the range of Knudsen number, gas flows can be categorized as three types:

- I. $Kn \ll 1$: Continuous flow
- II. $0.1 < Kn < 1$: Transition flow
- III. $1 \ll Kn$: Free molecular flow

For the flows in region I, physical quantities can be solved by a slip flow model, using Navier–Stokes equations with various velocity slip boundary conditions. The first expression of slip velocity can be traced back to the work of Maxwell. Burgdorfer (1959) applied the Maxwell's result for the lubrication problem, which was subsequently called the first-order slip velocity boundary condition, to derive a modified Reynolds equation. For the flows in regions II and III, the assumption of the gas as a continuum breaks down. Analyzing such flows requires a molecular-based model from the kinetic theory of gases, which describes the statistical properties of gas molecules. The most typical molecular-based model for lubrication problems is the Fukui–Kaneko model (Fukui and Kaneko 1988) based on the numerical solution of the linearized Boltzmann equation and the generalized Reynolds equation. Fukui and Kaneko showed that the molecular flow again can be decomposed into Poiseuille flow and Couette flow. The mass flow rate for Couette flow in the molecular flow regime is identical to the solution obtained from the Navier–Stokes equations, but the Poiseuille flow rate is



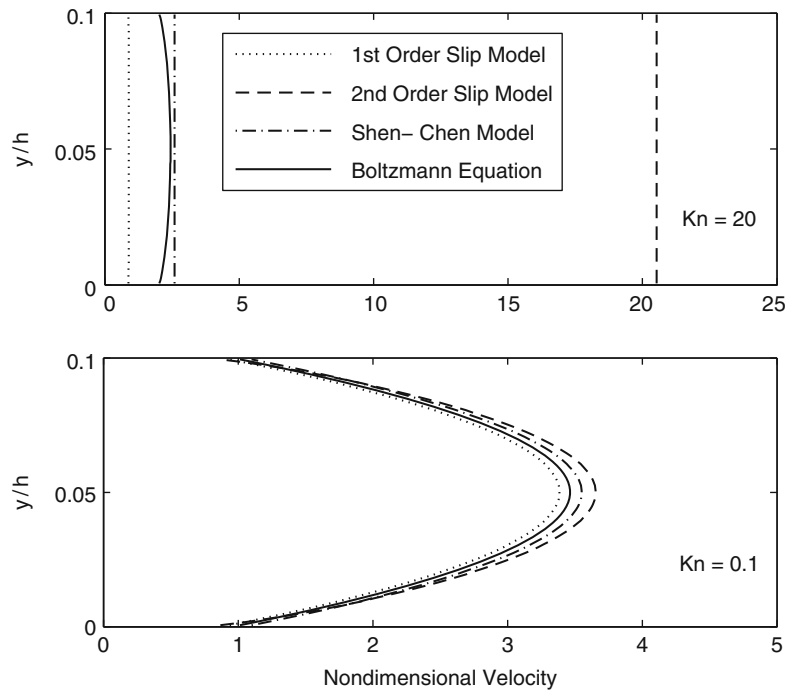
Molecular Gas Film Lubrication, Fig. 1 Poiseuille flow in two parallel plates



Molecular Gas Film Lubrication, Fig. 2 The nondimensional flow rates of Poiseuille flow as a function of Knudsen numbers

significantly altered due to molecular rarefaction. For Poiseuille flow, Fukui and Kaneko used the tabulated results of the numerical solution of the linearized Boltzmann equation obtained by Cercignani and Daneri (1963). The Fukui–Kaneko approach is believed to give the most accurate results for arbitrary Knudsen numbers among the existing compressible lubrication equations.

For Poiseuille flow in Fig. 1, the nondimensional Poiseuille flow rate predicted from the Boltzmann equation is plotted as a function of Knudsen number in Fig. 2 together with results from other slip models, assuming the diffuse reflection of gas molecules on the boundaries. The most prominent feature is the existence of a minimum (Cercignani and Daneri 1963), the so-called Knudsen minimum, around $Kn = 1$, which was confirmed by experimental results (Dong 1956). Compared to molecular-based models, slip models are simpler and



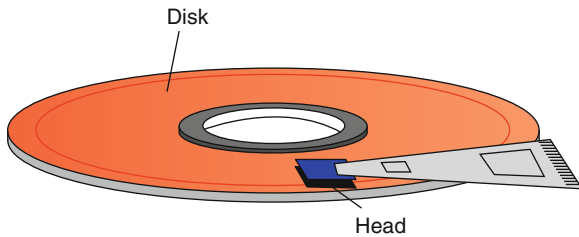
Molecular Gas Film Lubrication, Fig. 3 The nondimensional velocity profiles of Poiseuille flow with different Knudsen numbers

more efficient because they allow solving gas film lubrication problems using the continuum description. However, the first-slip model cannot be applied to large Knudsen numbers and does not predict the Knudsen minimum. Hsia and Domoto (1983) expanded the velocity slip at the boundaries to second-order, subsequently called second-order slip boundary condition, and proposed the corresponding Reynolds equation. The second-order slip model does demonstrate the existence of Knudsen minimum, but it largely overpredicts the flow rate at large Knudsen numbers. A 1.5-order slip model is also introduced to reduce the overprediction (Mitsuya 1993). Shen and Chen (Shen et al. 2007) proposed a new slip model based on the solution of the Boltzmann equation and momentum balance in the surface region, hence avoiding the mathematical expansion to second order. In Fig. 2, the new model shows a better agreement with the Boltzmann equation. The corresponding nondimensional velocity profiles are plotted in Fig. 3. It can be seen from Fig. 3 that the first-order and second-order slip models give consistent results with Boltzmann equation for a small Knudsen number, but large discrepancies exist for a large Knudsen number. In both cases ($Kn = 0.1$ and $Kn = 20$), the velocities of gas molecules on the boundaries are nonzero, which exhibits molecular rarefaction effects.

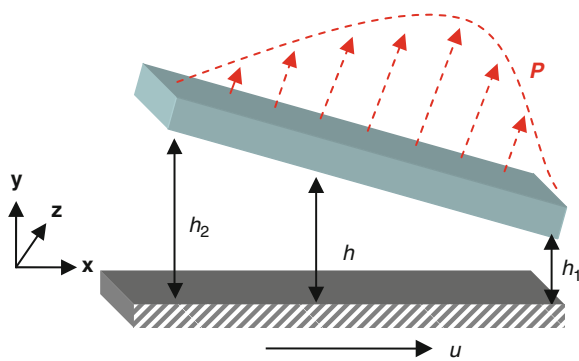
Although the solutions based on the linearized Boltzmann equation are generally considered most accurate for molecular gas film lubrication problems, there are limitations both in the solutions and the equation. First, the solutions of the Boltzmann equation are often based on the Krook approximation, i.e., relaxation time approximation, of the scattering term in the Boltzmann equation. A well-known deficiency of the Krook approximation is that the solutions obtained cannot lead to correct Prandtl number (Vincenti and Kruger 1965). Second, the Boltzmann equation assumes point particles and chaotic scattering processes (Chen 2005). In some molecular lubrication cases, molecular size can no longer be neglected. Remedies for such approximations include Monte Carlo simulation (Huang and Boggy 1997) and the Enskog model (Chapman and Cowling 1953).

Key Applications

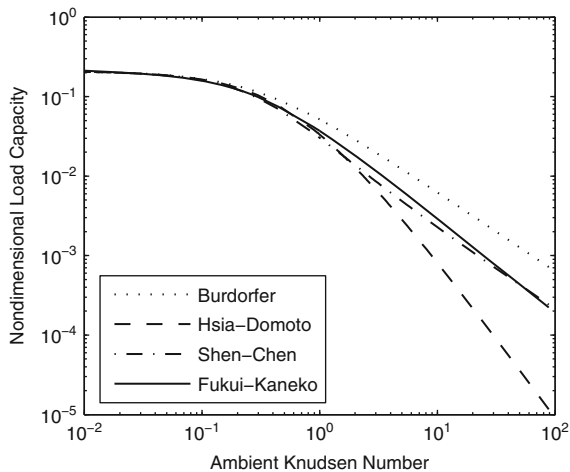
Molecular gas film lubrication has broad applications in MEMS devices (e.g., micromotors) and magnetic recording systems. For instance, when a hard disk drive works, its read/write head is “flying” just tens of nanometers above the surface of its magnetic disk, as shown in Fig. 4. The rapid spin of the disk creates a thin film of pressurized air, permitting the head to fly just above the disk and providing an extremely low friction at the head and



Molecular Gas Film Lubrication, Fig. 4 Schematic drawing of the read/write head and the magnetic disk in a hard disk drive



Molecular Gas Film Lubrication, Fig. 5 The inclined plane slider bearing



Molecular Gas Film Lubrication, Fig. 6 The nondimensional load capacity of different models

disk interface (HDI). The air bearing in the head-disk system of a hard disk drive is similar to the inclined plane slider bearing in Fig. 5. The flow inside the air bearing can be decomposed as two basic flows: Poiseuille

flow (pressure driven flow) and Couette flow (shear driven flow). If the head and disk surfaces are assumed to have the same feature of interacting with air molecules, the Couette flow rate is independent of the Knudsen number according to the symmetry of velocity profiles. Thus, in current HDI simulations of hard disk drives, molecular rarefaction effects are incorporated into the generalized Reynolds equation by correcting the Poiseuille flow rate. Figure 6 displays the nondimensional load capacity of the inclined plane slider as a function of the ambient Knudsen number for the case of the ratio $h_1/h_2 = 0.5$ and the bearing number $\Lambda = 10$, where the ambient Knudsen number is defined as the ratio of h_1 and the mean free path of air molecules at the ambient. The bearing number Λ is defined as $6\mu UL/(p_0 h_1^2)$, where μ is the viscosity of the gas, U is the disk rotation speed, L is the length of the slider, and p_0 is the ambient pressure. In comparison with the Fukui-Kaneko model, the first-order slip model overpredicts the load while the second-order slip model underpredicts the load.

Cross-References

- [Gas Bearing Applications](#)
- [Reynolds Equation for Compressible Fluid or Gas Film](#)

References

- B. Bhushan, *Handbook of Micro/Nano Tribology* (CRC Press, Boca Raton, FL, 1995)
- A. Burgdorfer, The influence of the molecular mean free path on the performance of hydrodynamic gas lubricated bearing. *ASME J Basic Eng* **81**, 94 (1959)
- C. Cercignani, A. Daneri, Flow of a rarefied gas between two parallel plates. *J. Appl. Phys.* **34**, 3509 (1963)
- S. Chapman, T.G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, Cambridge, UK, 1953)
- G. Chen, *Nanocale Energy Transport and Conversion* (Oxford, New York, 2005)
- W. Dong, in *Vacuum flow of gases through channels with circular, annular and rectangular cross sections*, University of California Radiation Laboratory, Report No. UCRL-3353 (1956)
- S. Fukui, R. Kaneko, Analysis of ultra-thin gas film lubrication based on linearized Boltzmann equation: First report-derivation of generalized lubrication equation including thermal creep flow. *ASME J Tribol* **110**, 253 (1988)
- B.J. Hamrock, S.R. Schmid, B.O. Jacobson, *Fundamentals of Fluid Film Lubrication* (Marcel Dekker, New York, 2004)
- Y.T. Hsia, G.A. Domoto, An experimental investigation of molecular rarefaction effects in gas lubricated bearings at ultra-low clearances. *ASME J. Lubrication Technol.* **105**, 120 (1983)
- W. Huang, D.B. Bogy, Three-dimensional direct simulation Monte Carlo method for slider air bearings. *Phys. Fluids* **9**, 1764 (1997)
- E.H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill, New York, 1938)
- Y. Mitsuya, Modified Reynolds equation for ultra-thin film gas lubrication using 1.5-order slip-flow model and considering surface accommodation coefficient. *ASME J. Tribol.* **115**, 289 (1993)

- J.W. Powell, A review of progress in gas lubrication. *Rev. Phys. Technol.* **1**, 96 (1970)
- S. Shen, G. Chen, R.M. Crone, M. Anaya-Dufresne, A kinetic-theory based first order slip boundary condition for gas flow. *Phys. Fluids* **19**, 085101 (2007)
- W.G. Vincenti, C.H. Kruger Jr., *Introduction to Physical Gas Dynamics* (Krieger, Florida, 1965)

Molten Salt

- [Ionic Liquid Lubricants](#)

Molten Salts Oxinitrocarburizing

- [Nitrocarburizing: Arcor\(R\), a Controlled Ionic Liquid Process](#)

Molybdenum Disulfide

- [MoS_x Coatings by Closed-Field Magnetron Sputtering](#)

Monolayer Adsorption

- [Chemical Vapor Deposition Processes for Boundary Lubricants](#)

Monotonic and Cyclic Deformation in Single Crystals

XIAO-WU LI
Institute of Materials Physics and Chemistry, College of Sciences, Northeastern University, Shenyang, People's Republic of China

Synonyms

CSS behavior – cyclic stress-strain behavior; LCF – low-cycle fatigue; PSB – persistent slip band

Definition

Monotonic and cyclic deformation in single crystals generally refers to the changes in properties that occur in monocrystalline material, resulting from the application of monotonic loads (e.g., tension or compression) and cyclic loads (e.g., cyclic straining or cyclic stressing), respectively. Cyclic plastic deformation is the most decisive phenomenon in the fatigue process of single crystals because it can cause irreversible changes in the dislocation microstructure of single crystals. There is no fatigue failure if no repeated plastic deformation is involved in the crystal.

Scientific Fundamentals

Monotonic Deformation in Single Crystals

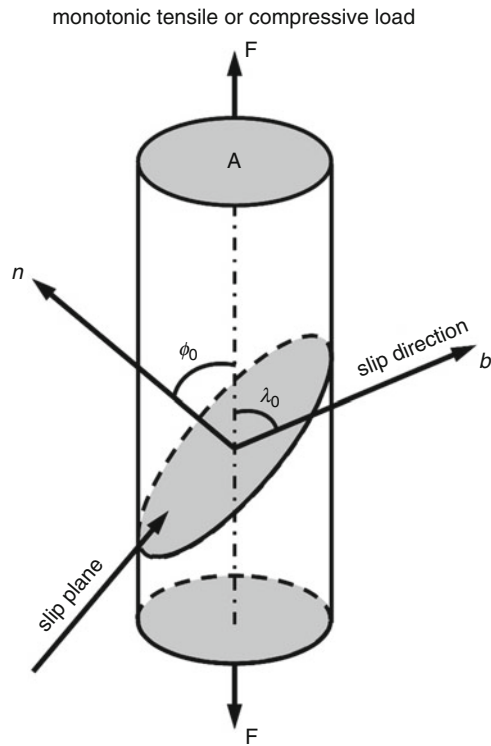
Schmid's Law

Plastic deformation of a single crystal subjected to monotonic stress is generally characterized in terms of resolved shear stress τ and resolved shear strain γ , which act on a specific slip plane along a specific slip direction. Such specific slip plane(s) and slip direction(s) consist of preferred slip system(s), over which crystallographic slip by dislocation motion governs the plastic deformation of single crystals. Slip directions are commonly the crystallographic directions with the shortest distance between atoms, and slip planes are usually densely packed planes. **Figure 1** schematically presents a diagram of a single crystal subjected to a monotonic tensile or compressive load F , showing the orientations of the preferred slip plane and slip direction. The slip plane normal (\mathbf{n}) and the slip direction (\mathbf{b}) are oriented at angles of ϕ_0 and λ_0 , respectively, with respect to the loading direction. The initial resolved shear stress τ_0 on the slip plane should be expressed by the resolved shear force, $F \cos \lambda_0$, along \mathbf{b} divided by the area of the slip plane, $A / \cos \phi_0$:

$$\tau_0 = \frac{F}{A} \cos \phi_0 \cos \lambda_0 = \sigma \cos \phi_0 \cos \lambda_0 = \Omega \sigma \quad (1)$$

where σ is the longitudinal stress and $\Omega = \cos \phi_0 \cos \lambda_0$ is the Schmid factor or orientation factor, which has a maximum value of 0.5 as the crystallographic orientation is satisfied with $\phi_0 = \lambda_0 = 45^\circ$ (e.g., $[\bar{1}49]$ orientation). The Schmid factor Ω also relates the increment of resolved shear strain $d\gamma$ on a slip system to the increment in axial strain $d\epsilon$. For small strain increments, $d\gamma = d\epsilon / \Omega$.

The prerequisite to the onset of plastic deformation is described by the well-known *Schmid's Law*, that is, if a crystal is stressed, slip begins just when the resolved shear stress τ_0 acting on a slip system attains a critical



Monotonic and Cyclic Deformation in Single Crystals, Fig. 1
Slip elements of a single crystal in monotonic tensile or compressive load

value τ_c , often called the critical resolved shear stress. This stress τ_c is dependent mainly on temperature, material purity, and strain rate; it increases with decreasing temperature and with increasing impurity content and strain rate: (1) Decreasing temperature means less thermal energy is available to overcome obstacles (e.g., Peierls stress, impurity, solute atoms, etc.), thus causing a need for a higher τ_c . (2) The higher impurity level leads to more short-range interactions of dislocations with those obstacles and thus to a higher τ_c . (3) Increasing strain rate means that a shorter period of time is available for a certain strain level, reducing the effect of thermal contributions in facilitating the movement of dislocations, namely the thermal assistance in moving dislocations is thus less pronounced, and therefore the τ_c is higher. A list of slip systems and values of τ_c for some typical metal single crystals at room temperature (Dieter 1976) is provided in Table 1.

Stress-Strain Behavior

A general description of the stress-strain behavior (or work-hardening behavior) of a single crystal is presented in the schematic Fig. 2. For an fcc (e.g., Cu, Ag, Ni) single crystal oriented for single slip, the curve of the shear stress τ versus the shear strain γ can be divided into three distinct stages. Stage I commences after initial elastic deformation, and the work-hardening rate immediately following yielding is low; this region is referred to as “easy glide,” which is generally characterized by primary

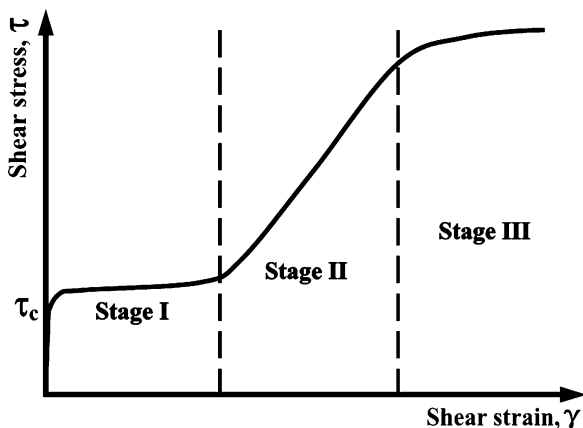
Monotonic and Cyclic Deformation in Single Crystals, Table 1 List of slip systems and values of the critical resolved shear stress τ_c for some typical metal single crystals at room temperature (Dieter 1976)

Crystal type	Metals	Purity (%)	Slip plane	Slip direction	τ_c (g/mm ²)
fcc	Cu	99.999	(111)	[110]	65
		99.98	(111)	[110]	94
	Ag	99.99	(111)	[110]	48
		99.97	(111)	[110]	73
		99.93	(111)	[110]	131
	Ni	99.8	(111)	[110]	580
bcc	Fe	99.96	(110), (112), (123)	[111]	2,800
	Mo		(110)	[111]	5,000
hcp	Zn	99.9999	(0001)	[1120]	18
	Cd	99.996	(0001)	[1120]	58
	Mg	99.996	(0001)	[1120]	77
	Ti	99.99	(1010)	[1120]	1,400
		99.9	(1010)	[1120]	9,190

fcc face centered cubic, bcc body-centered cubic, hcp hexagonal close-packed

slip and uniformly spaced slip lines. The transition from Stage I to Stage II behavior is associated with the onset of secondary slip and the attendant decrease in mean slip distance. The slope of the curve in Stage II, called “linear hardening” region, is large, indicating a significant increase in the work-hardening rate of the crystal. As the crystal deforms in Stage II, the dislocation density unceasingly increases and cross slip tends to occur, ultimately impelling the formation of dislocation cell structures. At a still higher stress level, which corresponds to Stage III, “exhaustion hardening” region, the work-hardening rate decreases in association with the onset of cross slip. The strain extent of “easy glide” region depends on crystal orientation, crystal perfection, and temperature. A higher degree of crystal perfection and a low temperature promote easier glide. As the crystal is oriented so that the resolved shear stress on other potential slip systems is rather low, the strain extent of easy glide is greater. Likewise, the extent of Stage II is reduced as temperature increases, since recovery processes operate more effectively at higher temperatures. In addition, because cross slip occurs more readily in crystals with high stacking fault energy (SFE), the transition from Stage II to Stage III takes place at lower stress levels in crystals having a high SFE, which usually exhibit pronounced Stage III deformation even at low applied stresses.

For hcp single crystals, the number of slip systems is comparatively limited, and the relevant Stage I in the stress-strain curve is often found to be longer. In contrast, in the case of bcc single crystals, the number of slip systems is comparatively greater, and multiple slip can be readily operated during deformation, so that Stage I in



Monotonic and Cyclic Deformation in Single Crystals, Fig. 2 Schematic of the typical shear stress–shear strain curve of an fcc single crystal

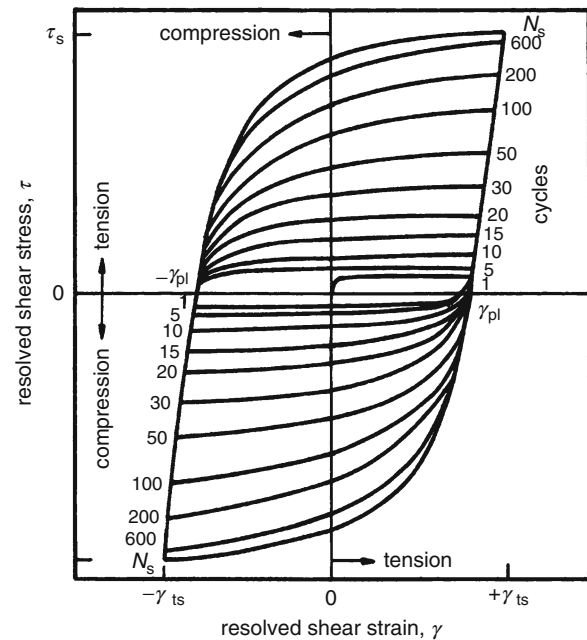
the stress-strain curve is usually hard to observe. For bcc and hcp single crystals, their stress-strain curves are also closely related to the crystal purity, strain rate, and temperature. Under a certain suitable condition, their stress-strain curves can also exhibit three obvious stages similarly, as shown in Fig. 2.

Cyclic Deformation in Single Crystals

The fatigue process of materials is conditioned by cyclic plastic deformation. The most conclusive findings of cyclic plastic deformation mechanisms in fatigue have been achieved mainly through studies on fcc single crystals, in particular Cu single crystals oriented for single slip. The main results have been well reviewed by several investigators (Basinski and Basinski 1992; Laird 1996; Suresh 1998; Wang et al. 1999). The most interesting points are focused upon the establishment of a cyclic stress-strain (CSS) curve, the observation of persistent slip band (PSB), and its corresponding dislocation structures.

Cyclic Hardening

A well-annealed fcc single crystal oriented for single slip will harden when it is subjected to cyclic deformation between constant plastic strain amplitude limits under



Monotonic and Cyclic Deformation in Single Crystals, Fig. 3 Schematic of typical variation in resolved shear stress τ as a function of the resolved shear strain γ for a single-slip-oriented fcc single crystal (Suresh 1998)

fully reversed loading at room temperature in air, as shown in Fig. 3. The peak stress increases rapidly at first and then more slowly approaches a value, the saturation stress τ_s , after N_s cycles. To characterize the globe fatigue damage under cyclic loads, the cumulative plastic strain, Γ , is frequently used for a nominal measure of damage accumulation, which is defined as (Suresh 1998)

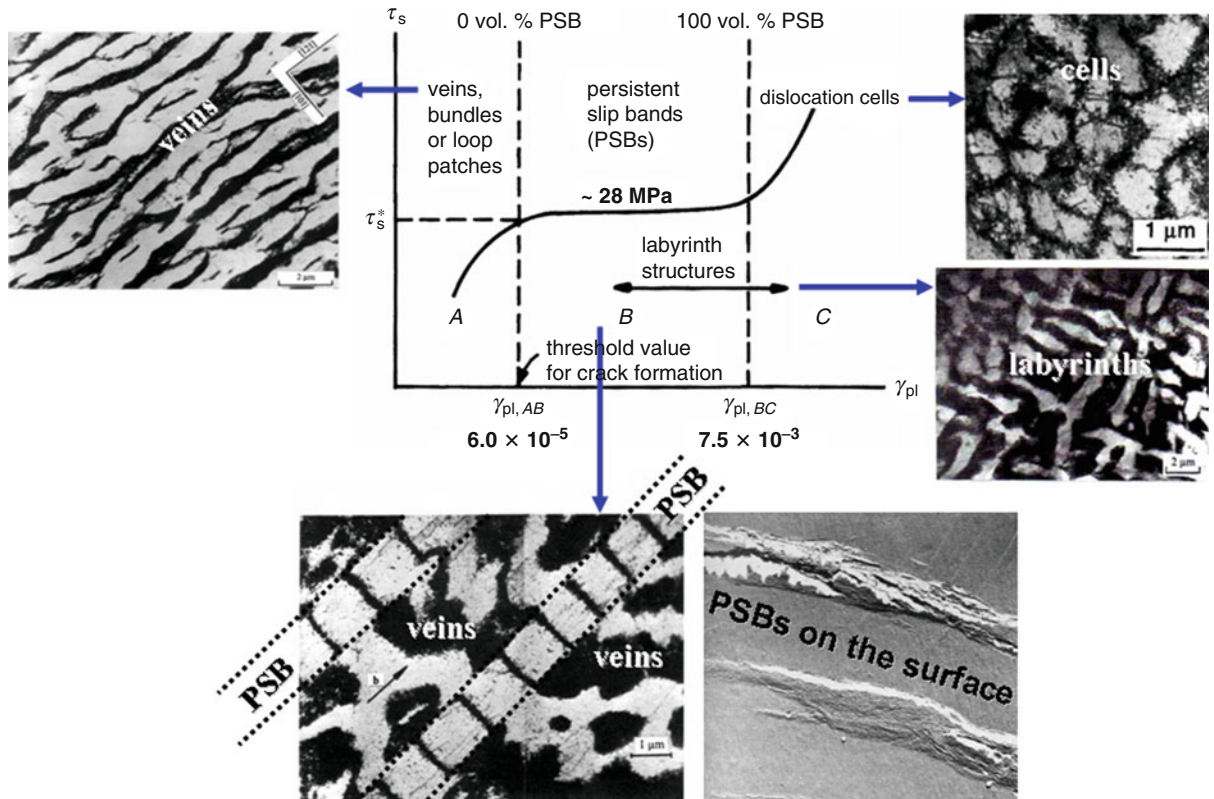
$$\Gamma = 4 \sum_{i=1}^N \gamma_{pl,i} \quad (2)$$

where $\gamma_{pl,i}$ is the resolved plastic shear strain amplitude in the i th cycle and N is the total number of cycles. For fully reversed cycling under a constant γ_{pl} , $\Gamma = 4N\gamma_{pl}$.

Cyclic Stress-Strain Curve

A plot of saturation stress τ_s versus the plastic strain amplitude γ_{pl} generates the cyclic stress-strain (CSS) curve. Mughrabi (1978) carried out constant plastic strain amplitude tests on single-slip-oriented Cu single crystals over a wide range of amplitudes and obtained a

well-documented CSS curve, which clearly demonstrates three regions marked by A, B, and C, as shown in Fig. 4. At low values of plastic strain amplitude ($\gamma_{pl} \leq \gamma_{pl,AB} = 6 \times 10^{-5}$), denoted region A, work hardening occurs during cyclic loading. Region A is followed by a strain-independent Region B, known as the plateau. The plateau region extends over two decades of plastic strain amplitudes ($6 \times 10^{-5} = \gamma_{pl,AB} \leq \gamma_{pl} \leq \gamma_{pl,BC} = 7.5 \times 10^{-3}$). A further increase in plastic strain amplitude ($\gamma_{pl} \geq \gamma_{pl,BC} = 7.5 \times 10^{-3}$) leads to an increase in the saturation stress, which is region C. One of the most visible features of cyclic saturation is the formation of the highly localized and largely reversible slip bands, PSBs, which often act as preferential sites for fatigue crack nucleation. The plateau region B is characterized by the formation of the first PSBs at its lower end ($\gamma_{pl,AB} = 6 \times 10^{-5}$) and by the complete filling of the crystal volume with PSBs at its upper end ($\gamma_{pl,BC} = 7.5 \times 10^{-3}$) (see Fig. 4). In the plateau region, τ_s remains essentially constant (~ 28 MPa) due to the localized cyclic deformation in PSBs. For single



Monotonic and Cyclic Deformation in Single Crystals, Fig. 4 Schematic of the CSS curve of Cu single crystals oriented for single slip showing three distinct regions as well as the corresponding features of dislocation structures. (Mughrabi et al. 1979; Basinski et al. 1980; Ackermann et al. 1984; Suresh 1998)

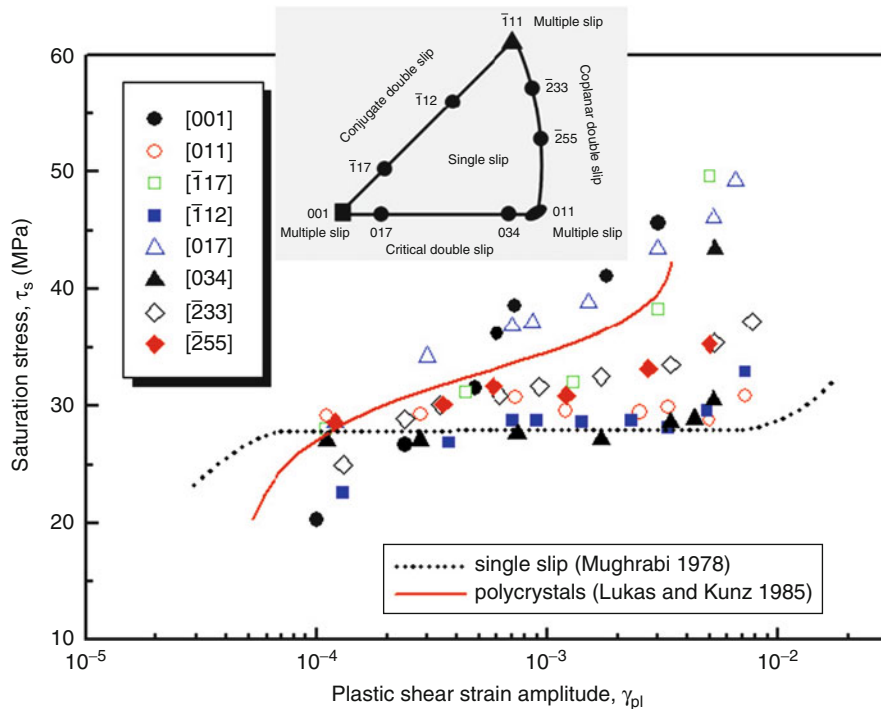
crystals of Cu, Ni, and Ag, cyclically deformed at room temperature, it is interesting to find that the ratio of the saturation plateau stress τ_s^* for PSB formation to the shear modulus (G) is approximately the same ($\sim 6.5 \times 10^{-4}$) (Mughrabi et al. 1979).

Dislocation Structures

As shown in Fig. 4, the dislocation structure in cyclically deformed Cu single crystals oriented for single slip strongly depends upon the applied γ_{pl} . In the plateau region, dislocation configuration can be described as two-phase structures (i.e., PSB ladders and matrix veins) (Winter 1974; Mughrabi et al. 1979). The ladder structure, corresponding to surface PSBs, is soft and carries most of the applied plastic strain, while the vein structure is hard and contributes little to cyclic deformation. The stress plateau behavior is accomplished by adjusting the amount of PSBs to correspond to the composed γ_{pl} . At γ_{pl} below the lower end of the plateau, the dislocation structure is occupied only by veins, whereas at γ_{pl} near the upper end of the plateau or beyond the plateau, labyrinth and cell structures that are associated with multiple slip may develop.

Orientation Effects on the CSS Behavior

Attempts have been made to select Cu single crystals as model materials to reveal the fatigue deformation mechanisms of polycrystalline materials. However, in actual polycrystalline materials, the orientation of grains is randomly distributed and many grains will be in all kinds of double- or multiple-slip orientations, especially in the regions around grain boundaries where double or multiple slip is frequently seen to activate due to elastic and plastic strain incompatibility between adjacent grains. Therefore, it may not be appropriate to relate simply the cyclic deformation behavior of single-slip-oriented crystals to that of polycrystals in which double or multiple slip is a general phenomenon. Obviously, the study of the cyclic deformation behavior of double- or multiple-slip-oriented single crystals is of particular importance for a better understanding of the cyclic deformation behavior in polycrystals. In recent years, the plateau behavior in the CSS curve and the cyclic strain hardening behavior, as well as the fatigue limits of differently oriented Cu single crystals, were surveyed by Li et al. (1999a, b, 2001a). Figure 5 presents the CSS curves of differently oriented Cu single crystals. It is clear that the CSS curves can be roughly



Monotonic and Cyclic Deformation in Single Crystals, Fig. 5 The CSS curves of differently oriented Cu single crystals. The CSS curves of single-slip-oriented Cu single crystals and Cu polycrystals are also included in the figure for comparison (Mughrabi 1978; Lukas and Kunz 1985; Li et al. 1999a)

divided into three groups according to the stress plateau behavior: (1) As with single-slip crystals (for which the CSS curve is hardly influenced by orientation), the CSS curve shows a clear plateau, such as [011], [034], and $\bar{1}12$ crystals. (2) There is no clear plateau, but a quasi-plateau in the CSS curve might exist within a certain range of strain amplitude γ_{pb} , such as $\bar{2}33$, $\bar{2}55$, and [017] crystals. (3) The shear saturation stress τ_s increases monotonically with increasing γ_{pb} ; there is no plateau in the CSS curve, such as [001] and $\bar{1}17$ crystals. The different plateau behaviors in single-, double-, and multiple-slip Cu crystals are summarized as follows.

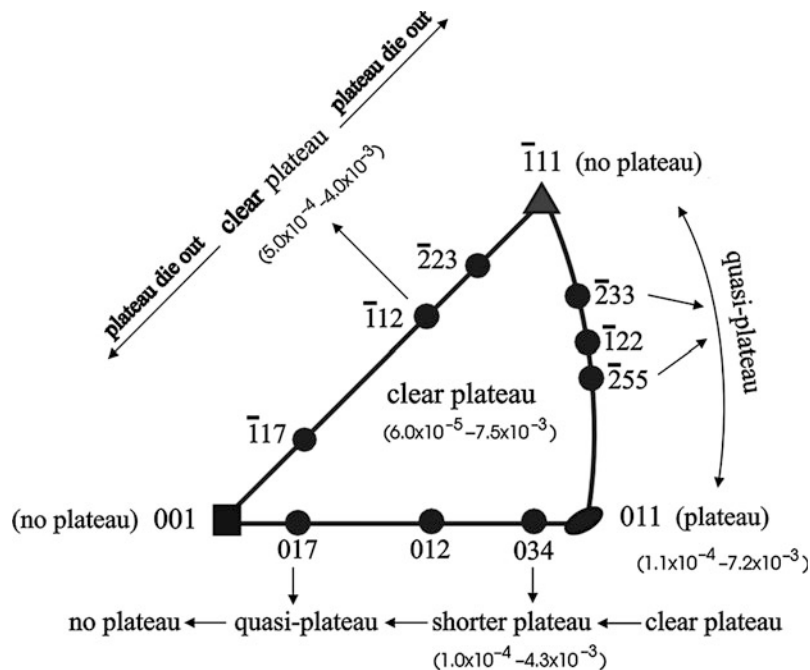
Orientations on the [011]– $\bar{1}11$ side: The CSS curves of the $\bar{2}33$ and $\bar{2}55$ crystals clearly show similar quasi-plateau regions. Consequently, for the coplanar double-slip-oriented crystals located on the [011]– $\bar{1}11$ side of the stereographic triangle, the transition from a clear plateau region in [011] crystal to no plateau region in $\bar{1}11$ crystal is linked through the presence of a quasi-plateau in between the two.

Orientations on the [001]–[011] side: The CSS curve of the [034] crystal exhibits a plateau in the strain range 1.0×10^{-4} – 4.3×10^{-3} , which is shorter than that of the [011] crystal. The CSS curve of the [017] crystal shows a narrow quasi-plateau over the strain range 5.0×10^{-4} to 1.5×10^{-3} . Therefore, when the crystal orientation varies

from [011] to [001] on the [001]–[011] side of the stereographic triangle, the plateau region in the CSS curve diminishes, that is, the CSS curve of the [011] exhibits a wide plateau region, the plateau region of the [034] crystal become shorter, the CSS curve of the [017] crystal shows a quasi-plateau over a narrow strain range, and the CSS curve of the [001] crystal does not show any plateau region. It seems that the CSS behavior varies from more like that of single-slip-oriented crystals to polycrystals when the crystallographic orientation changes from [011] to [001].

Orientations on the [001]– $\bar{1}11$ side: The CSS curve of the $\bar{1}12$ crystal exhibits a clear plateau region over the strain range 5.0×10^{-4} – 4.0×10^{-3} , which is shorter than that of the single-slip crystal, although the saturation shear stresses are nearly the same. However, the CSS curve of the $\bar{1}17$ crystal does not show any plateau region. Besides, a $\bar{2}23$ crystal has no clear saturation phenomenon, which is similar to that of a $\bar{1}11$ crystal. In summary, for the crystals located on the [001]– $\bar{1}11$ side of the stereographic triangle, the CSS curve of a $\bar{1}12$ crystal shows a clear plateau region, while the plateau regions of the CSS curves for crystals located on both sides of a $\bar{1}12$ crystal tend to die out.

The orientation dependence of the plateau behavior of Cu single crystals is summarized schematically in Fig. 6.

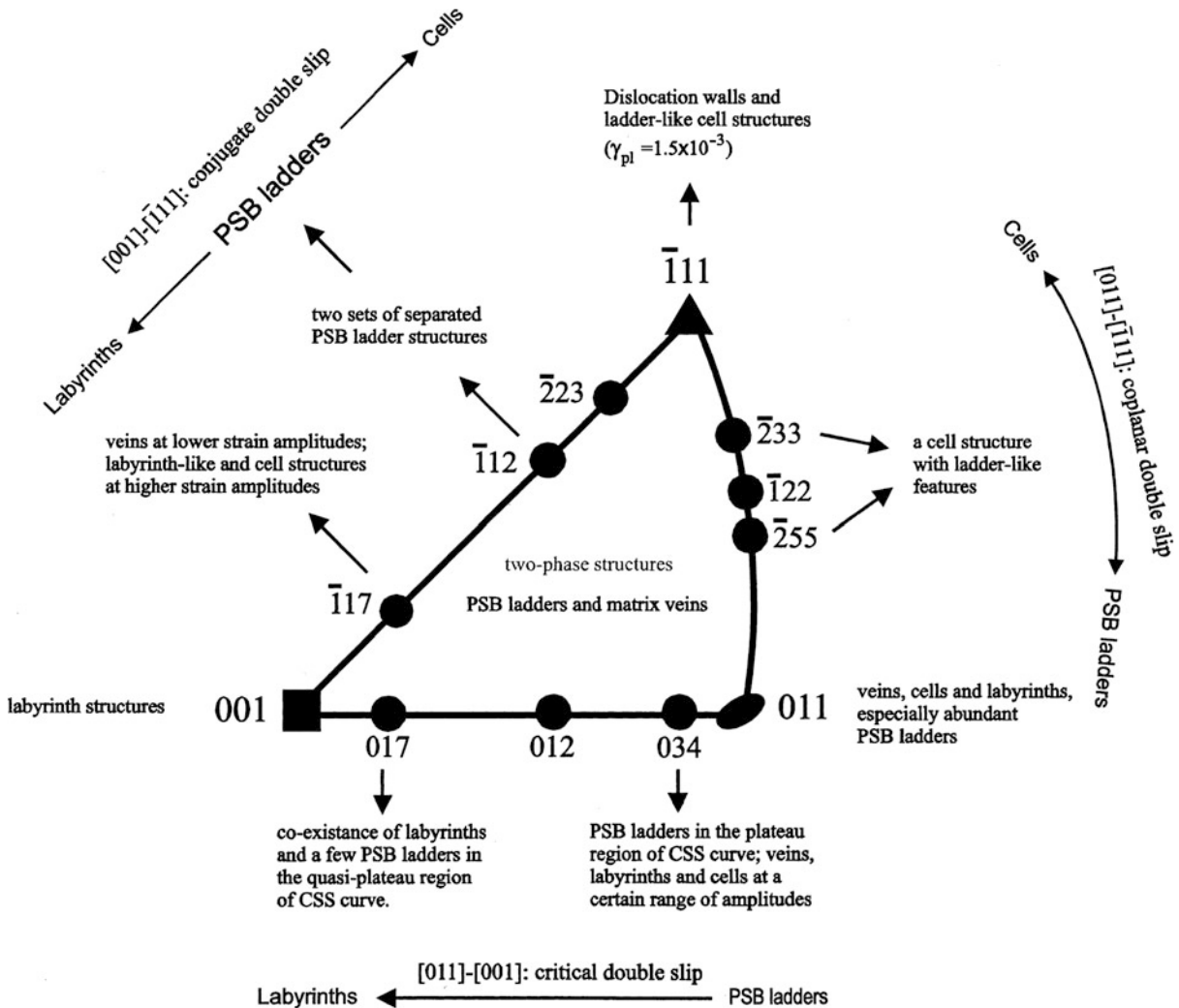


Monotonic and Cyclic Deformation in Single Crystals, Fig. 6 Standard stereographic triangle showing the different plateau behaviors in the CSS curve of differently oriented Cu single crystals (Li et al. 1999a)

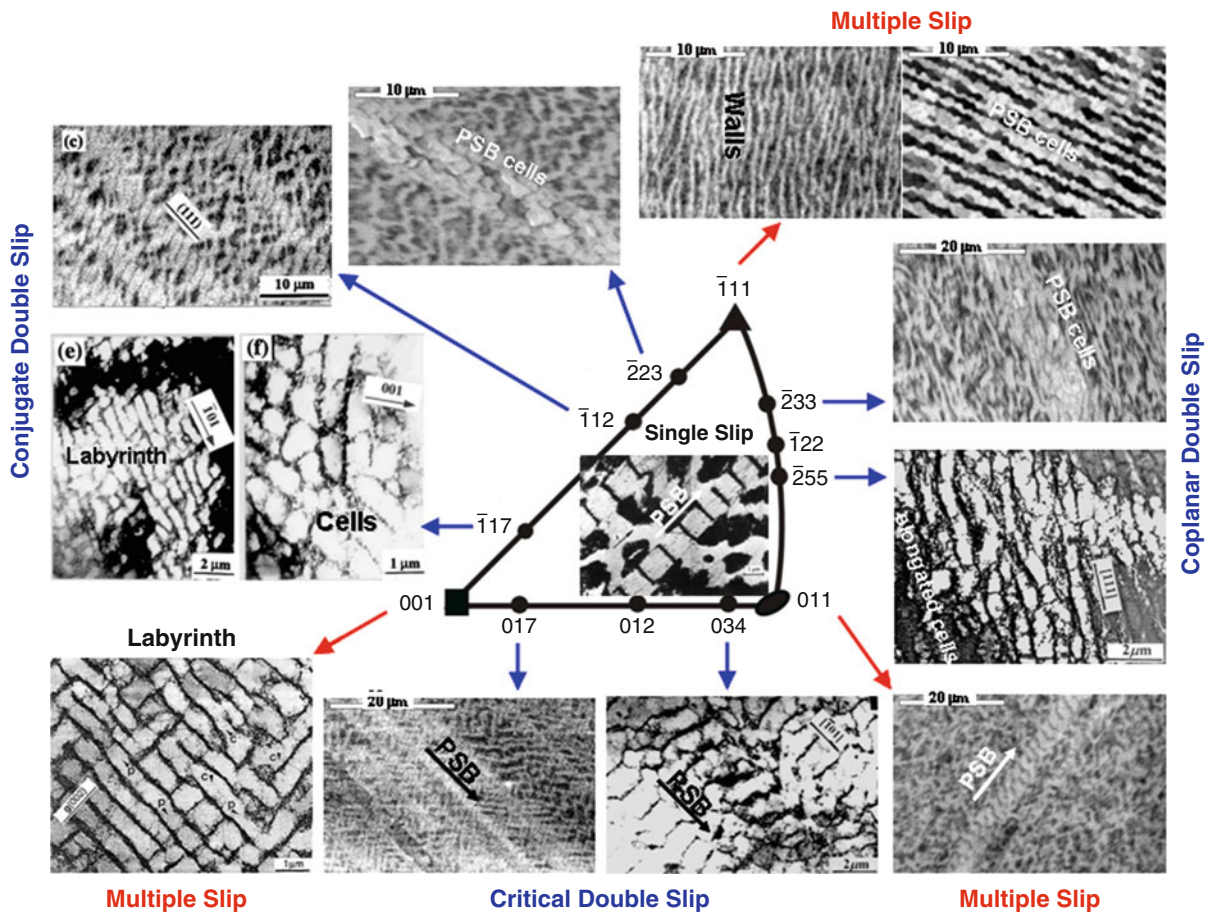
The existence or non-existence of a plateau in the CSS curves, as well as the corresponding plateau stress amplitude, depend not only on the modes and intensities of dislocation interactions among slip systems operating in the crystals, but also on the different slip deformation characteristics induced by the different geometrical relationships between orientations and corresponding slip systems in the differently oriented crystals. In general, the plateau region disappears only when multiple slip plays a determining role during cyclic deformation. Likewise, there also exists a relevant orientation dependence of the cyclic hardening behavior or the fatigue limits for Cu single crystals with various orientations (Li et al. 1999b, 2001a)

Orientation Effects on Dislocation Structures

The orientation dependence of cyclic dislocation structures of Cu single crystals is summarized schematically in Fig. 7 (Li et al. 2001b) and photographically in Fig. 8 (Mughrabi et al. 1979; Gong et al. 1996; Wang et al. 1997; Li et al. 2001b, 2002, 2009; Li and Zhou 2007). It can be clearly seen from Figs. 7 and 8 that the saturation dislocation structures change regularly with the variation in crystallographic orientation. For example, for the crystals located on the $[011]-[001]$ or $[011]-[\bar{1}11]$ side of the standard stereographic triangle, the dislocation structures change generally from PSB ladder structures to labyrinth structures or cell/wall structures, respectively, when the orientation varies from the $[011]$ orientation to the



Monotonic and Cyclic Deformation in Single Crystals, Fig. 7 Standard stereographic triangle showing the variation of dislocation structures with crystallographic orientation (Li et al. 2001b)



Monotonic and Cyclic Deformation in Single Crystals, Fig. 8 Typical dislocation structures of Cu single crystals with various crystallographic orientations at comparable plastic strain amplitudes (Mughrabi et al. 1979; Wang et al. 1997; Li et al. 2001b, 2002, 2009; Li and Zhou 2007)

$[001]$ orientation or to the $[\bar{1}11]$ orientation. However, for the crystals located on the $[001]$ – $[\bar{1}11]$ side, PSB ladder structures obviously form in the $[\bar{1}12]$ crystal, while PSB ladder structures for the crystals located on both sides of a $[\bar{1}12]$ crystal tend to die out and evolve as labyrinth structures or cell/wall structures when the orientation varies from the $[\bar{1}12]$ orientation to the $[001]$ orientation or to the $[\bar{1}11]$ orientation, respectively. Such an orientation dependence of cyclic dislocation structures provides a good account for the orientation dependence of the plateau behavior in the CSS curve of Cu single crystals (see Fig. 6).

Stacking Fault Energy Effects on Cyclic Deformation in Single Crystal

Although fundamental investigations of the cyclic deformation mechanism in some single crystals of wavy

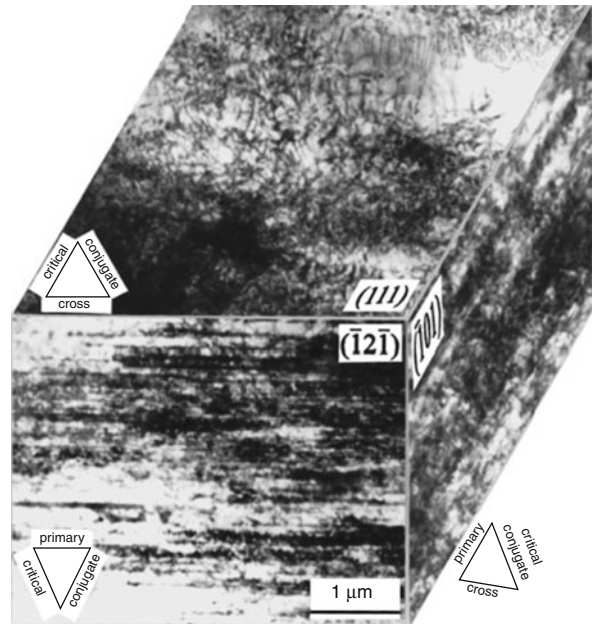
slip pure metal (e.g., Cu) have yielded a wealth of experimental and theoretical results, knowledge on the cyclic deformation mechanism of low stacking fault energy (SFE) or planar slip alloy single crystals still remains lacking. Li et al. (2003) summarized and studied the cyclic deformation behavior of some low-SFE alloy single crystals. Dilute Cu solid solution alloys behave with great similarity to their base metal, namely Cu, both in cyclic and dislocation behavior. However, with the increase in solute content, the SFE decreases and the change of slip mode from wavy slip to planar slip would thus take place, causing the cyclic behavior to become different from that of pure metals. For example, the ladder structures of PSBs, which can often be seen in cycled Cu single crystals, are not found in fatigued Cu-30 at.%Zn alloy single crystals, and the dislocation structures in cyclically deformed Cu-30wt.%Zn crystals oriented for single slip have two

basic configurations depending on the imposed strain amplitude, namely, dislocation segments and multipoles at low strain amplitudes and planar dislocation loops and tangles at high strain amplitudes. The cyclic hardening behavior of single-slip-oriented Cu-Al alloy single crystals with Al content less than 4 at.% is similar to that of Cu single crystals, whereas the crystals with Al content more than 11 at.% does not exhibit a clear cyclic saturation phenomenon. Persistent Lüder's bands (PLBs) consisting of densely dislocated arrays on closely spaced parallel primary slip planes act as the main form of strain localization in fatigued Cu-16 at.%Al single crystals, which is different from the case of Cu single crystals, where strain localization takes place in the form of PSBs featuring the ladder structure of dipolar walls.

To clarify the effect of the crystallographic orientation on the dislocation structures in low SFE materials as well as the role of slip deformation mode playing in the dislocation behavior during cyclic deformation of fcc crystals, the CSS behavior and dislocation structures of cyclically deformed Cu-16 at.%Al alloy single crystals oriented typically as $[\bar{1}23]$ for single slip and $[023]$ and $[\bar{1}17]$ for double slip have been studied and discussed in comparison with results of Cu single crystals [Li et al. 2003]. Unlike the case of Cu single crystals, the crystallographic orientation has almost no effect on the CSS behavior and corresponding dislocation structures of Cu-16 at.%Al alloy single crystals; this is true at least for the above three differently oriented crystals. The difference of cyclic dislocation structures between Cu and Cu-16 at.%Al crystals can be interpreted in terms of different slip deformation modes due to the addition of alloying element Al. The resultant dislocation structures of $[\bar{1}23]$, $[023]$, and $[\bar{1}17]$ Cu-16 at.%Al alloy single crystals present a typical planar morphology. In the primary slip plane, the dislocation structures consist mainly of multipolar arrays at low plastic strain amplitudes and develop into dislocation tangles at high plastic strain amplitude, whereas the dislocation structures in the foils normal to the primary slip plane change from low-density planar slip bands into well-developed PLBs with increasing plastic strain amplitude, as typically shown in Fig. 9. PLBs, where the applied plastic strains are highly localized, are expected to form in variously oriented Cu-16 at.%Al crystals because of the planarity of the dislocation structures.

Cyclic Deformation in BCC and HCP Single Crystals

Compared with the amount of productive research done on the cyclic deformation of fcc single crystals, less corresponding information is available to date on bcc and especially on hcp single crystals.



Monotonic and Cyclic Deformation in Single Crystals, Fig. 9
Three-dimensional views of the dislocation structures in cyclically deformed low-SFE $[023]$ double-slip-oriented Cu-16 at.%Al alloy single crystals at $\gamma_{pl} = 6.3 \times 10^{-3}$ (Li et al. 2003)

For single-slip bcc single crystals like α -Fe, Nb, and Mo, their CSS behavior is quite different from that of fcc crystals. The core of the dislocations, particularly for screw dislocations, in bcc crystals can be extended and rather complicated. Screw dislocations in $\langle 111 \rangle$ directions, for example, have a core with a threefold symmetry, leading to a basic asymmetry between the forward and backward movement of a dislocation. The particular nature of the screw dislocation core structure would induce a very high lattice friction (i.e., Peierls stress), so that bcc crystals usually exhibit strain-rate sensitivity, strong temperature dependence of cyclic deformation, and asymmetry of slip between tension and compression. Mughrabi et al. (1976, 1979) found that the CSS behavior of α -Fe single crystals is strongly dependent upon the plastic strain amplitude. At low plastic strain amplitude ($\leq 10^{-3}$), no cyclic hardening occurs and the cyclic strain is only accommodated by the motion of edge dislocations. In contrast, at high plastic strain amplitudes, cyclic deformation proceeds by the large-scale motion of edge and screw dislocations, and pronounced cyclic hardening as well as shape changes of the crystal are observed due to the asymmetric slip of screw dislocations under tension and compression. Ultimately, a cell structure forms. Unlike the case of Cu

single crystals, no PSBs have been found in fatigued α -Fe single crystals, regardless of the plastic strain amplitude. However, ill-defined slip bands, from which some cracks nucleate, are noticed.

Among hcp single crystals (e.g., Ti, Mg, Zr, etc.), the experimental research interests are mainly focused on the cyclic deformation in Ti single crystals. Tan et al. (1998) examined the cyclic deformation behavior of high-purity Ti single crystals with a wide range of orientations. The CSS behavior is found to fall into three groups, depending on orientation and the extent of slip and twinning. (1) For those crystals oriented for single prismatic slip, a stress plateau corresponding to PSB formation is observed in the CSS curves, and the plateau stress is 38 MPa. (2) For those crystals oriented for prismatic slip but for which cross-slip and twinning are favored, the plateau is suppressed and the flow stresses are higher. (3) For those crystals with orientations on the borders of the unit triangle, remarkable hardening occurs, and the CSS curves are steep. In this case, there are multiple cases of slip and twinning. The orientation-dependent microstructures in fatigued Ti crystals are observed. The $\{10\bar{1}0\} <11\bar{2}0>$ prismatic slip is the most easily operative slip mode during cyclic loading, except in the region near the orientation of $<0001>$, where $<c+a>$ pyramidal slips prevail. Double prismatic slip occurs for orientations near the edges of the standard triangle, and cross-slip on $(10\bar{1}1)$ occurs for orientations in the middle of the standard triangle. Cyclic twins are found in most of the fatigued single crystals, as typically shown in Fig. 10. The majority of the twins observed are in the types of $\{10\bar{1}2\}$, $\{11\bar{2}1\}$, and $\{11\bar{2}2\}$.

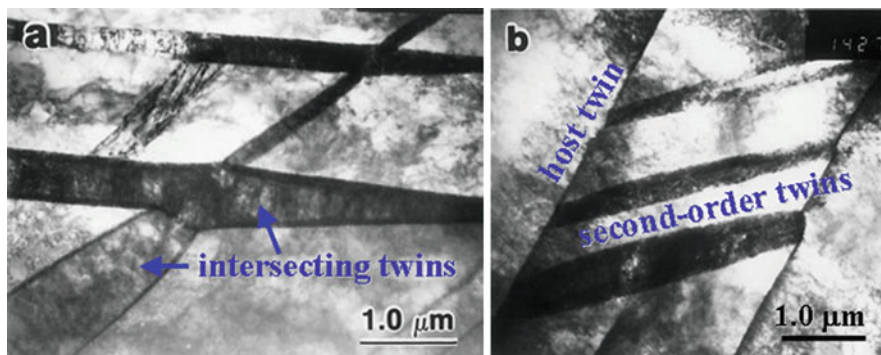
A comprehensive understanding of cyclic deformation behavior of bcc or hcp single crystals still requires further in-depth studies.

Key Applications

The whole fatigue damage process of a metal crystal, for the most general situation, can be classified into the following consecutive and partly overlapping stages:

1. Cyclic hardening and/or softening stage, which is characterized by the substructural and microstructural changes in the deformed crystalline material. Permanent damage has nucleated in this stage.
2. Microcrack creation, taking place just in a local area of the material, such as the surface layer.
3. Growth and coalescence of microcracks to form dominant macrocracks, which subsequently propagate in a stable manner. The governing factor of crack propagation is the localized cyclic plastic deformation within the plastic zone at the crack tip.
4. Structural instability and final failure.

Clearly, the nucleation of fatigue cracks results mainly from the irreversible microstructural changes caused by the cyclic plastic deformation. Therefore, understanding of microstructural changes induced by cyclic deformation is of extreme significance for microstructural design for fatigue resistance in engineering materials. However, on the one hand, for commercial materials, the complexities in initial microstructures, processing methods, impurity content, and so on often obstruct quantitative and even qualitative identification of cyclic deformation mechanisms. On the other hand, for polycrystalline materials, the existence in grain boundaries, triple junctions, and/or textures as well as the difference in grain size inevitably brings about a variety of difficulties in clarifying the fatigue mechanism. Accordingly, high-purity single crystals are often adopted as model materials for investigations (as described above), and these results arising from studies on single crystals lay a solid foundation for further



Monotonic and Cyclic Deformation in Single Crystals, Fig. 10 Cyclic deformation twins in fatigued Ti single crystals (Tan et al. 1998)

studying the cyclic deformation mechanism of polycrystalline materials. For example, the CSS curve of polycrystalline Ni has been well-predicted using a modified Taylor model, which takes into account experimental findings on relevant single crystals (Schwab and Holste 2002). Besides, knowledge of cyclic deformation behavior in the model material of single crystals greatly adds to our understanding of fatigue, and conclusive data obtained from such studies are fairly essential for microstructural design against fatigue in practical applications.

Knowledge of monotonic deformation behavior in single crystals is an indispensable key basis for the discussions of cyclic deformation mechanisms in single crystals.

Cross-References

- [Crystal Plasticity with Multiple Slip](#)
- [Cyclic Loading and Cyclic Stress](#)
- [Fatigue](#)
- [Fatigue Limit](#)

References

- F. Ackermann, L.P. Kubin, J. Lepinoux, H. Mughrabi, *Acta Metall.* **32**, 715 (1984)
- Z.S. Basinski, S.J. Basinski, *Prog. Mater. Sci.* **36**, 89 (1992)
- Z.S. Basinski, A.S. Korbel, S.J. Basinski, *Acta Metall.* **28**, 191 (1980)
- G.E. Dieter, *Mechanical Metallurgy* (McGraw-Hill, New York, 1976), p. 129
- B. Gong, Z.R. Wang, Z.G. Wang, Y.W. Zhang, *Mater. Sci. Eng. A* **210**, 94 (1996)
- C. Laird, *Physical Metallurgy*, eds. by R. W. Cahn, P. Hansen, vol. 3 (North Holland, 1996), p. 26
- X.W. Li, Y. Zhou, *J. Mater. Sci.* **42**, 4716 (2007)
- X.W. Li, Z.G. Wang, S.X. Li, *Phil. Mag. Lett.* **79**, 715 (1999a)
- X.W. Li, Z.G. Wang, S.X. Li, *Phil. Mag. Lett.* **79**, 869 (1999b)
- X.W. Li, Y. Umakoshi, Z.G. Wang, S.X. Li, *Phil. Mag. Lett.* **81**, 465 (2001a)
- X.W. Li, Z.F. Zhang, Z.G. Wang, S.X. Li, Y. Umakoshi, *Def. Diffus. Forum* **188–190**, 153 (2001b)
- X.W. Li, Y. Umakoshi, B. Gong, S.X. Li, Z.G. Wang, *Mater. Sci. Eng. A* **333**, 51 (2002)
- X.W. Li, X.M. Wu, Z.G. Wang, Y. Umakoshi, *Metall. Mater. Trans. A* **34**, 307 (2003)
- X.W. Li, Y. Zhou, W.W. Guo, G.P. Zhang, *Cryst. Res. Technol.* **44**, 315 (2009)
- P. Lukas, L. Kunz, *Mater. Sci. Eng.* **74**, L1 (1985)
- H. Mughrabi, *Mater. Sci. Eng.* **33**, 207 (1978)
- H. Mughrabi, C. Wüthrich, *Phil. Mag.* **A 33**, 963 (1976)
- H. Mughrabi, F. Ackermann, K. Herz, Persistent slip bands in fatigued face-centered and body-centered cubic metals, in *Fatigue Mechanisms*. Special Technical Publication, vol. **675** (American Society for Testing and Materials, Philadelphia, 1979), p. 69
- A. Schwab, C. Holste, *Acta Mater.* **50**, 289 (2002)
- S. Suresh, *Fatigue of Materials* (Cambridge University Press, Cambridge, 1998)
- X. Tan, H. Gu, C. Laird, N.D.H. Munroe, *Metall. Mater. Trans. A* **29**, 507 (1998); 513 (1998)

Z.R. Wang, B. Gong, Z.G. Wang, *Acta Mater.* **45**, 1379 (1997)

Z.G. Wang, X.W. Li, Z.F. Zhang, W.P. Jia, S.X. Li, *J. Mater. Sci. Technol.* **15**, 489 (1999)

A.T. Winter, *Phil. Mag.* **29**, 719 (1974)

MoS₂ and MoS_x

- [MoS_x Coatings by Closed-Field Magnetron Sputtering](#)

MoS₂ Coatings

- [Doped MoS₂ Coatings and Their Tribology](#)

MoS_x Coatings by Closed-Field Magnetron Sputtering

JULFIKAR HAIDER

School of Computing, Engineering and Information Sciences, Northumbria University, Newcastle upon Tyne, UK

Synonyms

[Molybdenum disulfide](#); [MoS₂ and MoS_x](#); [Solid lubricants based on MoS_x](#)

Definition

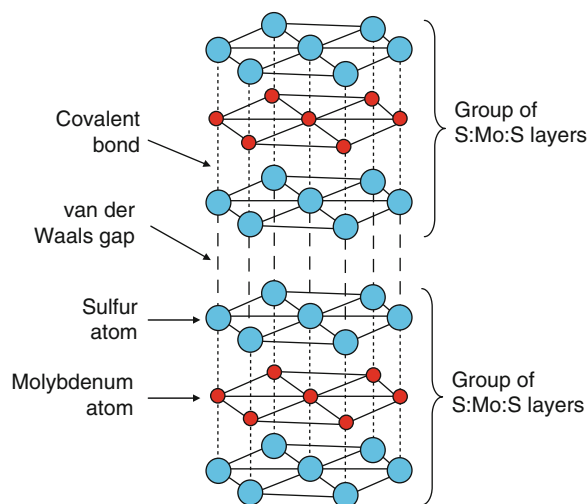
A MoS_x coating is described as a layer of solid lubricant material deposited (a few micrometers) onto wear components in order to enhance their tribological performance (e.g., lower coefficient of friction). MoS_x is the most appropriate solid lubricant coating in tribological applications, where the traditional liquid lubrication is not suitable (e.g., devices used in space) and where there is a need to fully or partially eliminate liquid lubricant (e.g., machining). Although MoS₂ can be used in the form of a dry powder, a fine dispersion in oil or grease, or impregnated into the matrix of porous materials to reduce friction, it is most frequently applied as a thin coating on components by the sputtering technique. As sputtered MoS₂ coatings are generally non-stoichiometric, with a deficiency of sulfur, it is termed as MoS_x, where x is approximately 1.8. Sputtered MoS_x coatings are an attractive choice due to their low coefficient of friction

(usually in the range of 0.01–0.25) and their ability to retain their favorable tribological properties in extreme environments (cryogenic temperature up to a maximum temperature of 300–400°C in air and up to 800°C in high vacuum). Sputtered MoS_x coatings obtained at optimum deposition conditions can display the lowest coefficient of friction (0.002) among the currently known solid materials under favorable operating conditions (i.e., ultra-high vacuum) (Spalvins 1992). A MoS_x coating is also regarded as a true solid lubricant as it can form a sacrificial transfer layer between two mating surfaces, which helps to reduce friction and wear. Fundamental knowledge, historical development, lubrication mechanism, deposition methods, and applications of MoS_x coatings have been reviewed by many authors (Lansdown 1999; Erdemir 2001; Holmberg and Matthews 2008).

Scientific Fundamentals

Basic Structure of MoS₂

MoS₂ has a hexagonal lamellar structure, similar to graphite, with individual layers of molybdenum (Mo) and sulfur (S) atoms. In a group of S:Mo:S layers, one Mo layer is surrounded by two S layers, each having six atoms, as illustrated in Fig. 1. The forces holding the atoms together in each group of S:Mo:S layers are strong covalent bonds, whereas the forces between the sulfur layers from two adjacent groups of S:Mo:S layers are weak van der Waals-type bonds. Thus, the combination of strong structure in two dimensions and weak structure in the third



MoS_x Coatings by Closed-Field Magnetron Sputtering,
Fig. 1 Schematic diagram of crystallographic structure in MoS₂

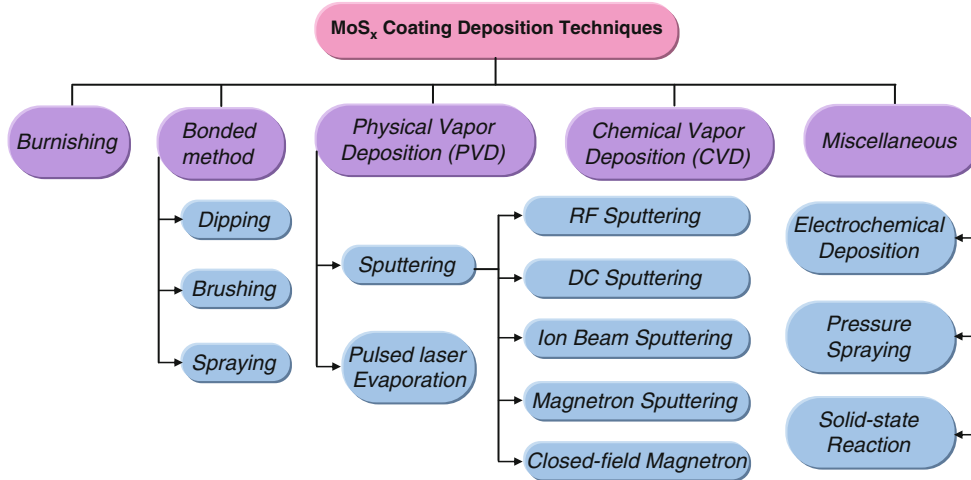
dimension makes MoS₂ a highly anisotropic material. The weak bonds allow the adjacent layers of sulfur atoms to slide easily over one another, which results in a low shear strength along the plane of the layers commonly known as “basal plane.” Thus, the low shear strength in MoS₂ structure forms the basis of low-friction characteristics. In order to obtain effective lubrication, the basal planes in a MoS_x coating should lie parallel to the sliding direction.

Deposition Methods for MoS_x Coatings

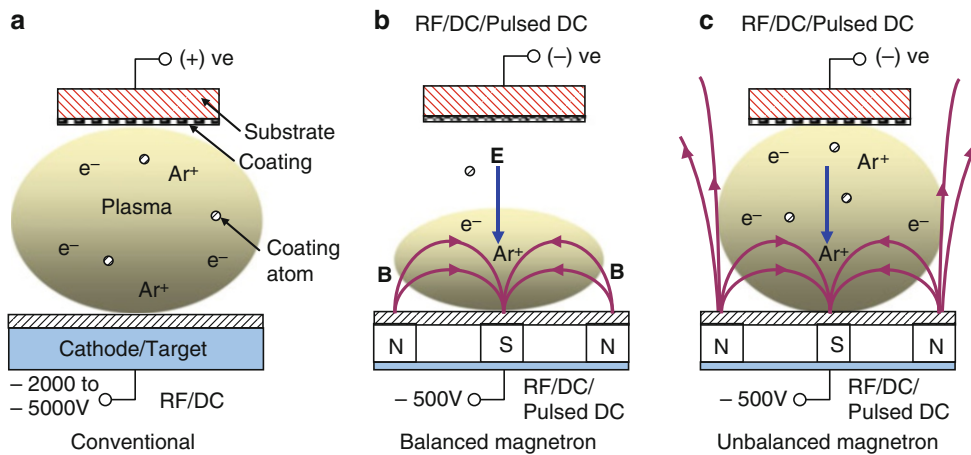
Various deposition techniques have been used to prepare MoS_x coatings, including burnishing, the bonded method, physical vapor deposition (PVD), and chemical vapor deposition (CVD), as shown in Fig. 2. Burnishing is the easiest and cheapest way to form a MoS₂ coating by rubbing MoS₂ powder onto a surface, but the coating provides a very short life. In the bonded method, MoS₂ powder mixed with a suitable binding agent is applied onto a component by dipping, brushing, or spraying. Although bonded MoS₂ coatings are cheaper to produce, their poor adhesion, non-uniform thickness, and relatively high porosity limit their wide use. With PVD, in particular, the sputtering process is most widely used for depositing MoS_x coatings. Many variants of the sputtering process are available, depending on the mode of power supply (DC or RF) or source of bombarding ions (magnetron or ion beam). MoS_x coatings can be deposited either by sputtering a MoS₂ target or Mo target in reactive gas (e.g., H₂S) environment. The CVD method has also been used to deposit a MoS_x coating on complex objects with good adhesion and without requiring any high vacuum. However, CVD is not a very attractive choice for MoS_x coating deposition due to some major shortcomings such as higher deposition temperature to form crystalline MoS_x coating (400–600°C), unwanted impurities in the coating from the metallorganic precursors, and dependence of the structural orientation on the precursors. In recent years, ion beam and laser beam-assisted deposition techniques have been developed to deposit good-quality MoS_x coatings. Other specialized techniques such as electrochemical deposition, pressure spraying, and solid-state reaction are not frequently reported.

Sputtering: A Method of Coating Deposition

In a basic sputtering system, coating materials (target/cathode) are mechanically dislodged by an energetic ion from a glow discharge plasma and subsequently deposited on a substrate placed nearby (Fig. 3a). The conventional sputtering process is limited in many ways: low deposition rate, low ionization efficiency, higher operating voltage (–2,000 to –5,000 V), higher operating pressure (>1 Pa),



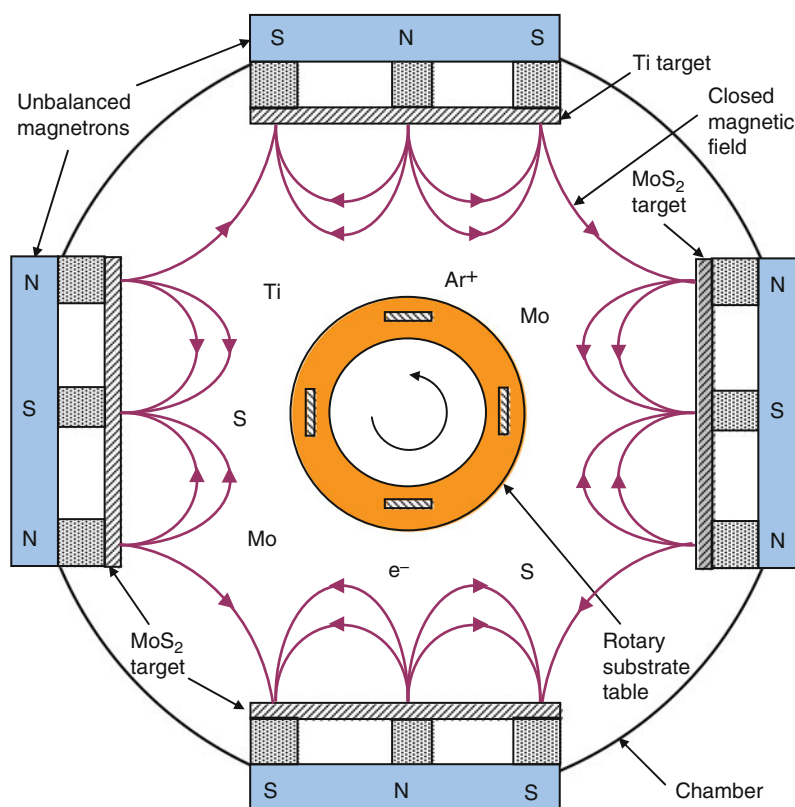
MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 2 Processes used for depositing MoS_x coatings



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 3 Different configurations of sputtering technology: (a) Conventional (b) Balanced magnetron, and (c) Unbalanced magnetron

and higher substrate heating. The development of magnetron sputtering (balanced) has overcome a majority of the problems associated with conventional sputtering. Electrons trapped by crossed electric and magnetic field ($\mathbf{E} \times \mathbf{B}$) in magnetron sputtering increase the ionization efficiency, which increases the sputtering rate, hence a higher deposition rate (Fig. 3b). However, low ion current density at the substrate ($\ll 1 \text{ mA/cm}^2$) due to the strong confinement of plasma near the target region does not contribute to the modification of coating structure. A later modification is the unbalanced magnetron sputtering system, where the outer magnetic assembly is made much stronger than the inner assembly in order to

flow some magnetic field lines towards the substrate (Fig. 3c). These additional magnetic field lines prevent ionization electrons from escaping the magnetron discharge and result in higher ionization around the negatively biased substrate. Therefore, the substrate receives higher ion current, which is favorable for improving adhesion and density of the coating. Quite recently, the emergence of a closed-field unbalanced magnetron sputtering (CFUBMS) system has even further increased the ionization level due to electron trapping by the closed magnetic field (Fig. 4). Deposition rate in CFUBMS is approximately ten times higher than conventional sputtering, and deposition can take place at much lower



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 4 Cross-section of a closed-field unbalanced magnetron sputtering (CFUBMS) chamber

pressure ($\sim 10^{-1}$ Pa) and voltage (~ 500 V). Further enhancement of the coating structure, density, and adhesion can be realized with the state-of-the-art CFUBMS system, owing to the substantial increase of ion current density drawn at the substrate (~ 100 times higher than balanced magnetron). The application of a pulsed power supply to the magnetron sputtering further improves the coating structure and process stability. An excellent review on the development of sputtering technology is presented in (Kelly and Arnell 2000).

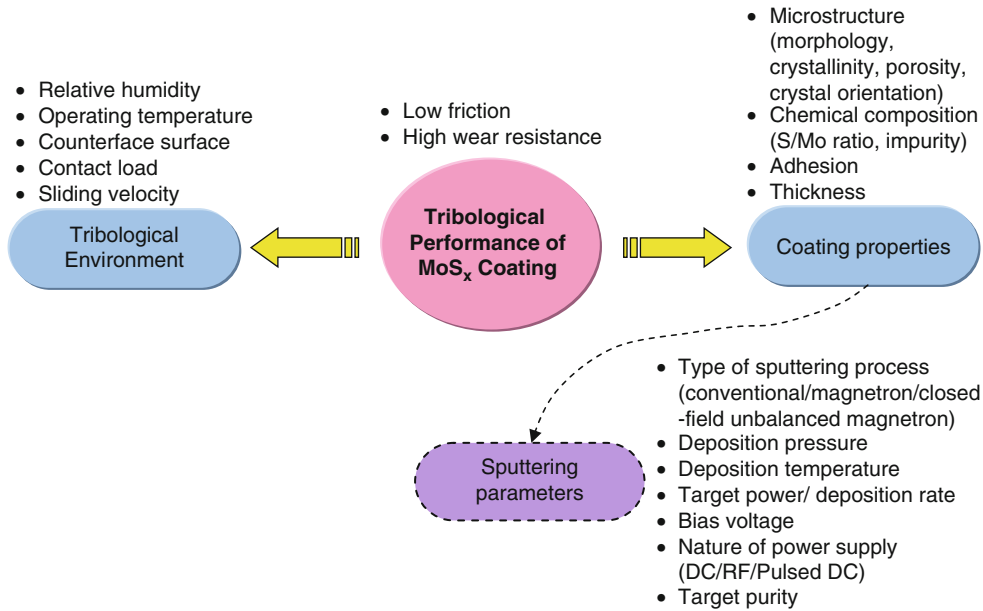
Sputtered MoS_x Coatings: Deposition, Characteristics, and Performance

Pioneering work by Spalvins during the late 1960s has established that sputtering is a remarkable technique for depositing high-quality MoS_x coating (Spalvins 1987). Since then, the sputtering process has been extensively investigated and accepted as the standard method of MoS_x coating production by the surface engineering community. The conventional DC or RF sputtering method was frequently used to produce MoS_x coatings

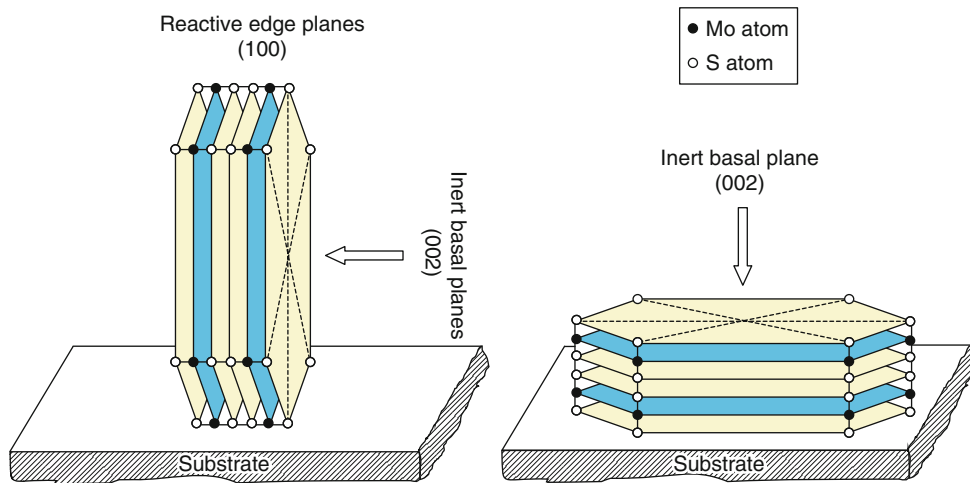
with low-friction coefficient and wear rate, but low deposition rate and a tendency to form porous columnar structure are cited as the major drawbacks. The advent of closed-field unbalanced magnetron sputtering (CFUBMS) has led to significant enhancement of coating properties (e.g., structural, chemical) as a result of improved deposition conditions. The tribological performance of MoS_x coatings is strongly dependent on tribological environment and coating properties, which are controlled by sputtering conditions, as shown in Fig. 5.

Coating Properties

In general, MoS_x coatings deposited by sputtering technique exhibit a crystalline structure, which is favorable for low-friction properties. However, a purely amorphous or quasi-amorphous (i.e., poorly crystallized) coating can also result, depending on the sputtering conditions. The brittle and abrasive nature of the amorphous coating is the reason why no lubricating properties with a high coefficient of friction (>0.4) are shown, but in some cases the quasi-amorphous coating can display good lubrication



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 5 Factors affecting the performance of MoS_x coatings



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 6 Schematic diagram of orientations of MoS₂ crystallites

properties. Surface morphology of MoS_x coatings gives an indication of the coating structure. Three distinctive surface morphologies are observed in MoS_x coatings with decreasing argon pressure during deposition: (a) needle-shaped, indicating a porous structure; (b) dome-shaped, indicating a dense structure; and (c) featureless, indicating a highly dense structure (Lauwerens et al. 2000).

Two types of crystallite orientations are generally observed in sputtered MoS_x coatings: basal planes

perpendicular to the substrate (Type I) and (2) basal planes parallel to the substrate (Type II), as shown in Fig. 6 (Spalvins 1992). A Type I coating is characterized by needle-like topography, high shear strength, and loosely packed columnar structure. Furthermore, the exposed reactive edge sites (100) easily oxidize in a humid environment, leading to higher coefficient of friction and poor wear resistance. On the other hand, a Type II coating having a basal plane (002) parallel to

the sliding direction can provide low shear strength and, hence, low friction and wear result. In addition, the protected edge sites in Type II coatings can offer higher resistance to oxidation. Type II coatings can be characterized by cluster shape topography and dense non-columnar structure.

MoS_x coatings produced by a conventional sputtering process (DC or RF) exhibit porous columnar structure with the basal plane oriented perpendicular to the substrate, which is typical of Type I coatings. This porous columnar structure is very prone to oxidation, and major part of the coating is worn away during sliding after breaking from the base of the column. Therefore, a very thin residual layer ($\sim 0.2 \mu\text{m}$) left on the substrate provides effective lubrication with a relatively shorter coating life. On the other hand, MoS_x coatings deposited by closed-field unbalanced magnetron sputtering exhibit long life, low coefficient of friction, and high resistance to humidity due to a good combination of dense non-columnar microstructure, favorable basal plane orientation, and excellent adhesion to the substrate. Post-deposition processes such as ion beam or laser beam treatment can also modify the structure of MoS_x coatings, making them more resistant to moisture. The reported thickness of MoS_x sputtered coatings usually ranges from 0.1 to 2.0 μm . Some degree of thickness reduction can be expected to take place during sliding, particularly in coatings with a porous, columnar structure. Spalvins (1992) pointed out that only a very thin coating of about 0.2 μm is required for effective lubrication. Coating properties do not vary considerably, with thickness within the range from 0.15 to 3.0 μm (Lauwerens et al. 2000).

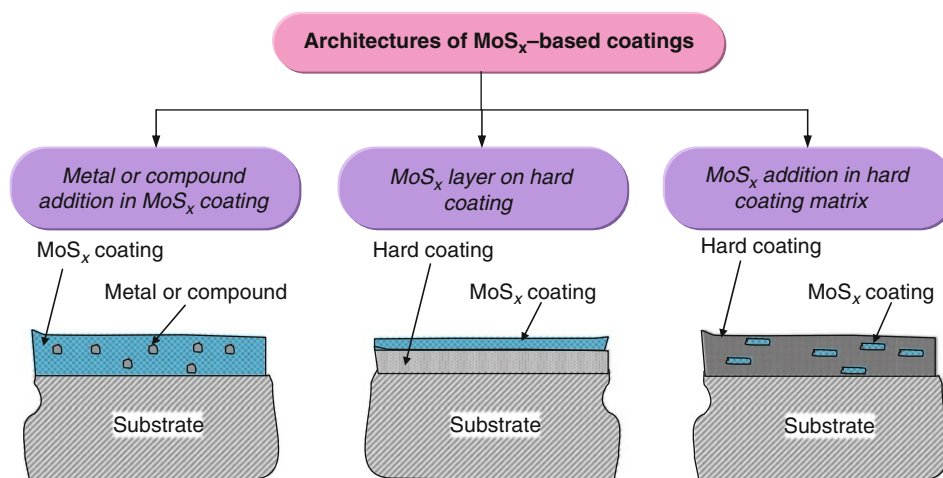
MoS_x coatings with a higher ratio of basal plane to edge plane (002)/(100) yield lower friction and wear rate and exhibit increased resistance to degradation in reactive environments. Ion bombardment on the growing coating caused by a negative bias voltage (-50 to -100 V) or unbalanced mode of the magnetron favors the formation of basal plane orientation (002) and suppresses the edge orientation to some extent. The opposite effect can be expected with a positive bias. Higher ion energy generated due to the pulsed power supply during deposition ($>175 \text{ eV}$) also promotes (002) orientation and suppresses (100) crystal growth. Hence, coatings deposited with pulsed DC exhibit much lower wear at high humidity compared with that made with DC power (Lauwerens et al. 2000). It has been demonstrated that maximum (002)/(100) is obtained at an optimum deposition rate (Muratore and Voevodin 2009) or lower cathode current. Deposition temperature and pressure can also influence the orientation of the crystal planes. MoS_x coatings with

higher (002)/(100) can be obtained by decreasing the deposition pressure and temperature in a magnetron sputtering process (Zhang et al. 2001). However, in some cases, it has also been seen that as-deposited coating with edge planes can be reoriented to basal planes during sliding (Fleischauer et al. 1989).

The low-friction properties of MoS_x coatings are strongly dependent on the stoichiometry ($\text{S/Mo ratio} = x$), which can vary from low to high sulfur content ($0.8 < x < 2.2$). However, an optimum combination of wear and friction properties can be obtained with a deficiency of sulfur in the coating (x values in the range of 1.2–1.8). The content of crystalline structure increases with increasing x value, which will improve the lubrication property of MoS_x coatings, and a minimum x value of 1.1 is required to maintain the hexagonal crystal structure in the coating (Spalvins 1992). Sulfur deficiency in MoS_x coatings is generally due to resputtering of sulfur atoms from the growing coating by the energetic ion bombardment under low pressure, high negative bias voltage, or strong unbalanced condition of the magnetron. Sulfur deficiency can also occur due to the substitution of sulfur atoms by oxygen atoms present in the chamber in a poor vacuum condition. MoS_x coatings can be contaminated by residual gases (e.g., H_2O , O_2) or other contaminants (e.g., S, C) from the previous coating cycle present in the deposition chamber. These contaminants subsequently affect crystalline orientation, morphology, coefficient of friction, and wear life of MoS_x coatings. Generally, the contamination level varies from 5% to 15%. Oxygen can react with MoS_x, forming either oxidation product (MoO_3 , which is detrimental to coating performance) or a new phase by substituting S atoms ($\text{MoS}_{2-x}\text{O}_x$, which is believed to be good for lubrication and coating endurance). Partial pressure of H_2O during deposition affects the microstructure and morphology of the as-deposited coating. With an increasing level of H_2O content in the sputtering chamber, MoS_x structure changes from Type II to Type I and finally to amorphous (Buck 1987).

Strong adherence of MoS_x coatings is very important for coating life, as adhesive failure of the coating leads to a short and unpredictable life. Sputter etching of the substrate surface through ion bombardment and depositing a thin metal layer prior to coating deposition enhance the adhesion of MoS_x coating. Surface pre-treatment also increases the density and chemical reactivity of nucleation sites, resulting in an increase in packing density (i.e., less porosity) of the columns in the coating (Fleischauer et al. 1989).

In contrast to the sputtering process, MoS_x coatings grown by ion beam-assisted deposition exhibit better



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 7 Classification of MoS_x-based coating architectures

tribological properties, which could be attributed to the improvement in microstructure, composition, and adhesion as a result of controlled ion bombardment. The pulsed laser method is also capable of producing MoS_x coatings with basal plane orientation parallel to the substrate, good stoichiometry, uniform thickness, and minimal contamination by background gas.

Tribological Environment

Operating conditions and environment have a profound impact on the durability and friction properties of MoS_x coatings. Under ultra-high vacuum and dry nitrogen environments, MoS_x coatings exhibit super-lubricant properties, with coefficient of friction as low as 0.002 (Donnet et al. 1996). The super-low friction properties of MoS_x coatings in vacuum can be related to the specific structural properties (e.g., stoichiometricity, crystallinity, free from contaminants), development of transfer film on the mating counter face, and friction-induced basal plane orientation during sliding, leading to very low interfacial shear strength. The friction coefficient increased to 0.013–0.15 and 0.2 when the testing environment was switched to a high vacuum or atmospheric environment (40% relative humidity), respectively. It has also been demonstrated that friction coefficient and wear rate increase with the increasing relative humidity level during wear testing. However, the rate of increasing friction coefficient and wear with the relative humidity is much higher in random/edge-oriented MoS_x coatings than basal-oriented MoS_x coatings, due to the fact that random-oriented MoS_x coatings oxidize more easily (Zhang et al. 2001). It should be noted that coatings with dense structure have only a shallow humidity-affected

zone that does not spread throughout the coating thickness, and the characteristics of friction and wear behavior are recoverable when the environment is shifted from high to low humidity conditions during wear testing. Coefficient of friction decreases with increasing contact load, possibly due to the orientation of the crystal structure in the sliding direction. However, higher load and sliding velocity will accelerate the wear of the coating.

Advances in MoS_x-Based Coatings

Ideally, in tribological applications, the solid lubricant coatings should not only exhibit low coefficient of friction but also high durability and environmental stability. Although MoS_x coatings exhibit excellent low friction properties in vacuum or inert gas environments, they struggle to provide high wear life due to their low hardness and susceptibility to oxidation in terrestrial environments. To address these issues, current research is concentrating on developing MoS_x-based coatings using three different approaches, as shown in Fig. 7.

Metal or Compound Additions in MoS_x Coatings

Enormous efforts have been made in improving the tribological properties of pure MoS_x coatings by incorporating metals (e.g., Au, Ni, Pb, Ti, Cr) or compounds (e.g., PbO, TiN, SbO₃) into the coating in the form of a composite or multilayer structure (Hilton et al. 1998; Simmonds et al. 1998; Renevier et al. 2001). Metals or compounds are generally co-sputtered with MoS_x on a rotating substrate to produce this coating architecture. It has been evidenced that the improvement comes through structural modification such as formation of

a densified structure by the suppression of columnar growth, tribologically favored basal-oriented structure, higher elastic recovery, higher hardness, and preferential oxidation of metal protecting the MoS_x from oxidation during wear. Optimum tribological properties of MoS_x coatings depend on the correct concentration metal/compound, varying from 5% to 20%. Ti is most frequently used as an inclusion in MoS_x coatings due to its good adhesive properties and compatibility with MoS₂.

More recently, Teer Coatings Ltd. in the UK has developed a composite MoS₂/Ti coating (MoSTTM) with maximum 18% Ti content by the closed-field unbalanced magnetron sputtering process. This coating exhibits high hardness (1,000–2,000 VHN); high density; extremely low coefficient of friction (0.02–0.08), depending on the load and atmospheric condition (humidity); remarkably low wear rate, even at high humidity; excellent adhesion; and higher operating temperature (450°C). It is believed that inclusion of a Ti atom causes the lattice distortion of MoS₂ structure, which increases the hardness and resistance to humidity of the coating. Figure 8 shows the schematic diagram of a conventional sputtered MoS_x coating with a porous, columnar structure and a MoST coating with a dense, non-columnar structure.

Another technique for improving the performance of MoS_x coatings is to produce a MoS_x/metal multilayer structure by adding a metal or compound in the form of repeating layers, with the thickness of individual layers in the nanometer scale (Simmonds et al. 1998). By optimizing the layer thickness, a reduction of coefficient of friction by 60% and an improvement in sliding lifetime by four to six times can be achieved with MoS_x/metal (Pb or Ti) nanoscaled multilayer coating compared with pure MoS_x coating in moist air. MoS_x/Ti interfaces promoted the growth of (002) basal plane, and the degree of (002) basal orientation was dependent on the number of MoS_x/metal interfaces. The metal-doped MoS_x coating has shown potential improvement over pure

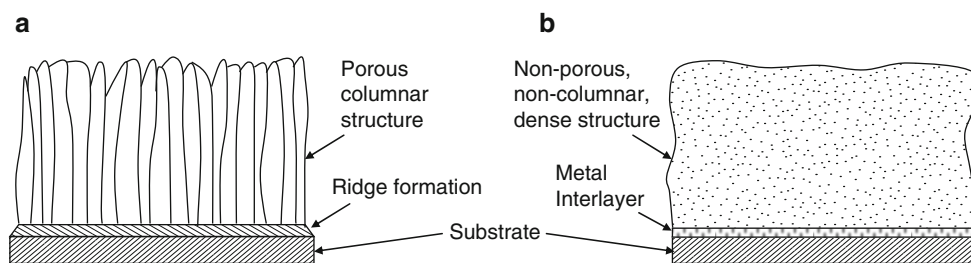
MoS_x, but lifetime is still limited owing to the soft lubricating matrix.

MoS_x Layer on Hard Coating

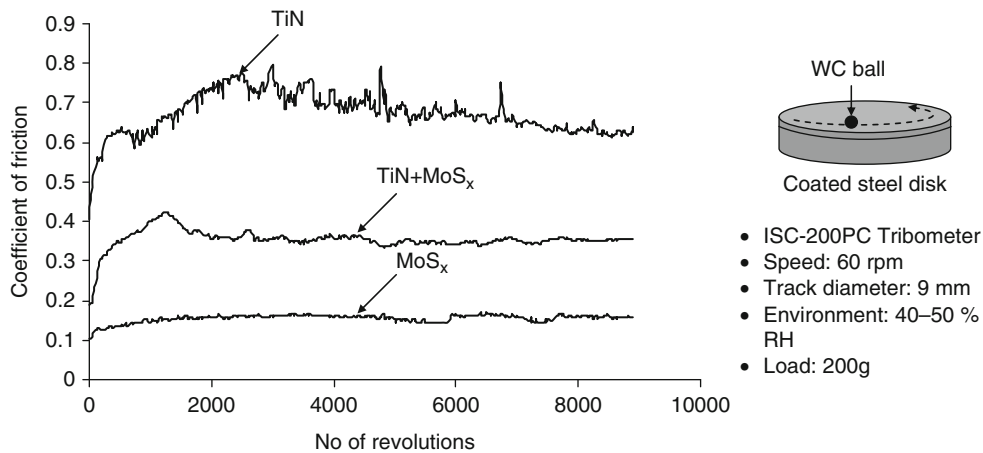
In this architecture, the MoS_x coating is deposited on top of a hard coating (e.g., TiN, CrN, TiCN, TiAlN, DLC, Si₃N₄) (Xu et al. 1999; Renevier et al. 2001). This double-layer approach can enhance the performance of the coating in tribological applications by the combined effect of a friction-reducing MoS_x top layer and a hard and load-bearing under-layer. Layered composite coatings of TiN/MoS₂ deposited by the magnetron sputtering technique have shown better fretting-resistant performance than single-layer TiN or MoS₂ under the same testing conditions. The solid lubrication property of MoS₂ coatings could be fully developed by a strong support from the TiN coating underneath (Xu et al. 1999). The hard coating layer in this architecture, produced either in the form of a multilayer or graded structure, and the soft layer on top of that, either as pure MoS_x or MoS_x/Ti, have proved successful in practical machining applications such as milling by extending tool life and easy chip flow (Renevier et al. 2001). Although soft MoS_x coatings can be removed very quickly, the lubrication effect can still be maintained for extended periods, possibly due to the small amount of MoS_x that remains trapped within the surface morphology of the hard coating.

MoS_x Addition in Hard Coating Matrix

Recently, combined hard-solid lubricant coatings have been produced by adding MoS_x into a hard coating matrix (e.g., TiN, TiAlN, TiSiN, TiB₂, CrN, CrB₂) with the intent of improving the frictional characteristics of the hard coatings (Gilmore et al. 1998; Haider et al. 2005; Gangopadhyay et al. 2009). These coatings can provide both high wear resistance and low friction coming from the hard and soft parts, respectively. MoS_x could be present in the hard coating, either as a separate phase or as



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 8 (a) Pure MoS_x coating deposited by conventional sputtering and (b) MoS₂/Ti (MoST) coating deposited by closed-field magnetron sputtering



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 9 Coefficient of friction of different coated samples obtained by pin-on-disk wear test

elements, but both are equally effective in obtaining the lubrication effect. Again, the low-friction properties can be maintained throughout the entire lifetime of the coating as MoS_x is uniformly distributed across the whole thickness of the coating. The concentration of the MoS_x needs to be properly controlled in order to obtain maximum lubrication without considerably sacrificing the hardness. MoS_x concentrations from 7% to 11% are suggested to achieve optimum tribological properties. A TiN-MoS₂ coating produced by unbalanced magnetron sputtering has demonstrated high hardness, exceeding 20 GPa, and friction coefficients of about 0.1 in normal atmospheric conditions. In a separate investigation, a TiN+ MoS_x coating deposited by closed-field magnetron sputtering demonstrated a significant reduction of coefficient of friction and wear rate compared with pure TiN coating, as shown in Figs. 9 and 10, respectively.

Key Applications

Traditionally, MoS_x coatings are used in space applications. Significant improvements in the properties and durability of MoS_x coatings as a consequence of technological advancement in the sputtering process and innovative coating designs have even spanned its applications in terrestrial environments.

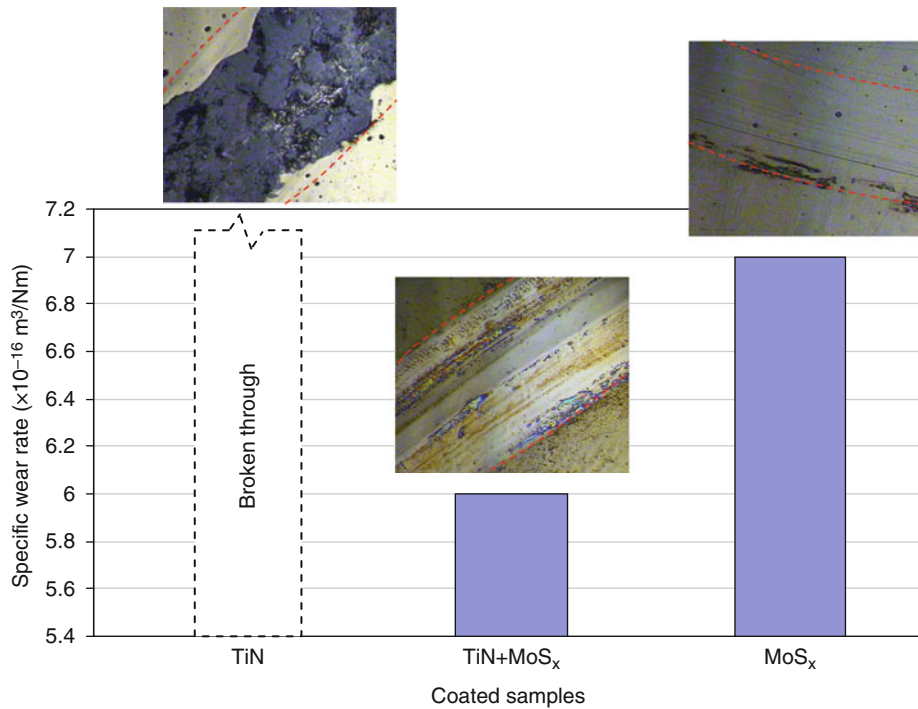
Space/Vacuum Technology

MoS_x coatings have attracted a great deal of attention in space/vacuum applications due to their ability to deliver excellent lubrication under extreme conditions such as high vacuum, high and low temperatures, and

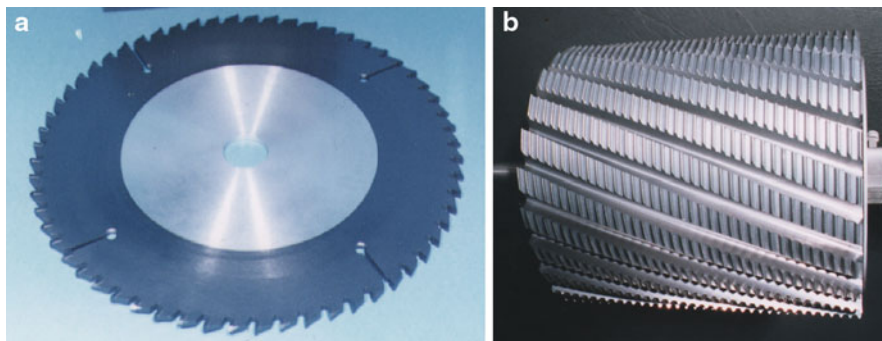
space radiation. Some examples of the application of sputtered MoS_x coatings in spacecraft and satellite are solar array drives, slip rings, antenna pointing, control systems, despin mechanisms, and rack and pinion gears' swivel mechanisms.

Cutting and Forming Tools

Lately, interest has been growing in the coating industries to develop low-friction MoS_x-based coatings for dry machining applications. MoS_x-based coatings, either as a top layer on hard coatings (e.g., TiN, TiAlN) (Renevier et al. 2000) or in a composite form (e.g., TiN-MoS₂) (Gilmore et al. 1998), have found applications in milling, drilling, punching, and hobbing. Generally, two to five times improvement in tool life has been achieved with these coatings in dry machining, depending on the applications. Commercial coaters such as Teer Coatings Ltd., UK (www.teercoatings.co.uk) and Guhring, Germany (<http://www.guhring.com>) have already developed their own brands for the MoS_x-based coatings (MoSTTM and MolyGlide[®], respectively) deposited by magnetron sputtering. Figure 11 shows a MoST-coated circular saw and hob for machining applications. The MoS_x-based coatings allow improved chip flow with a lowered coefficient of friction and reduced cutting force, heat generation, and built-up edge formation in the cutting tool. Hence, improved tool performance, higher productivity, and better workpiece quality result. MoS_x coatings have also found application as releasing material in plastic moulds, extrusion dies, and stamping tools owing to their excellent anti-adhesive or nonsticky properties.



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 10 Comparison of wear rates among TiN, TiN + MoS_x, and MoS_x coatings in a pin-on-disk wear test (test conditions are the same as shown in Fig. 9)



MoS_x Coatings by Closed-Field Magnetron Sputtering, Fig. 11 MoST (MoS₂/Ti) coated (a) circular saw and (b) hob (Courtesy: Teer Coatings Ltd., UK; www.teercoatings.co.uk; with permission)

Automotive Industry

In recent years, attention has been focused on improving the tribological characteristics of moving parts in automotive engines owing to the increasing demand for high power, high speed, improved fuel economy, and less environmental pollution. MoS_x-based solid lubricant coating (MoST) has shown promising results in improving the service life of automotive parts such as fuel injection systems, pistons, piston rings.

Cross-References

- [Adaptive Hard Coatings Design Based on the Concept of Self-organization During Friction](#)
- [D.C. Sputtering](#)

References

- V. Buck, Preparation and properties of different types of sputtered MoS₂ films. *Wear* **114**, 263–274 (1987)

- C. Donnet, J.M. Martin, T.L. Mogne, M. Belin, Super-low friction of MoS₂ coatings in various environments. *Tribol. Int.* **29**, 123–128 (1996)
- A. Erdemir, Solid lubricants and self-lubricating films, in *Modern Tribology Handbook*, ed. by B. Bhushan, 2001st edn. (CRC Press, Boca Raton, 2001), pp. 787–818
- P.D. Fleischauer, M.R. Hilton, R. Bauer, Effects of microstructure and adhesion on performance of sputter-deposited MoS₂, solid lubricant coatings, in *Mechanics of Coatings*, ed. by D. Dowson, C.M. Taylor, M. Godet. Tribology Series, vol. 17 (Elsevier, Amsterdam, 1989), pp. 121–128
- S. Gangopadhyay, R. Acharya, A.K. Chattopadhyay, S. Paul, Composition and structure–property relationship of low friction, wear resistant TiN–MoS_x composite coating deposited by pulsed closed-field unbalanced magnetron sputtering. *Surf. Coat. Technol.* **203**, 1565–1572 (2009)
- R. Gilmore, M.A. Baker, P.N. Gibson, W. Gissler, M. Stoiber, P. Losbichler, C. Mitterer, Low-friction TiN–MoS₂ coatings produced by dc magnetron co-deposition. *Surf. Coat. Technol.* **108–109**, 345–351 (1998)
- J. Haider, M. Rahman, B. Corcoran, M.S.J. Hashmi, Deposition and characterization of hard-solid lubricant coating by closed-field magnetron sputtering. *Surf. Coat. Technol.* **200**, 1080–1083 (2005)
- M.R. Hilton, G. Jayaram, L.D. Marks, Microstructure of cosputter-deposited metal- and oxide-MoS₂ solid lubricant films. *J. Mater. Res.* **13**, 1022–1032 (1998)
- K. Holmberg, A. Matthews, *Coatings Tribology — Properties, Techniques and Applications in Surface Engineering*. Tribology and Interface Engineering Series, vol. 28 (Elsevier, Amsterdam, 2008), pp. 211–225
- P.J. Kelly, R.D. Arnell, Magnetron sputtering: a review of recent developments and applications. *Vacuum* **56**, 159–172 (2000)
- A.R. Lansdown, in *Molybdenum Disulphide Lubrication*, ed. by D. Dowson. Tribology Series, vol. 35 (Elsevier, Amsterdam, 1999)
- W. Lauwerens, J. Wang, J. Navratil, E. Wieers, J. D’haen, L.M. Stals, J.P. Celis, Y. Bruynseraede, Humidity resistant MoS_x films prepared by pulsed magnetron sputtering. *Surf. Coat. Technol.* **131**, 216–221 (2000)
- C. Muratore, A.A. Voevodin, Control of molybdenum disulfide basal plane orientation during coating growth in pulsed magnetron sputtering discharges. *Thin Solid Films* **517**, 5605–5610 (2009)
- N.M. Renevier, N. Lobiondo, V.C. Fox, D.G. Teer, J. Hampshire, Performance of MoS₂/metal composite coatings used for dry machining and other industrial applications. *Surf. Coat. Technol.* **123**, 84–91 (2000)
- N.M. Renevier, J. Hampshire, V.C. Fox, J. Witts, T. Allen, D.G. Teer, Advantages of using self-lubricating, hard, wear-resistant MoS₂-based coatings. *Surf. Coat. Technol.* **142–144**, 67–77 (2001)
- M.C. Simmonds, A. Savan, H. Van Swygenhoven, E. Pflüger, S. Mikhailov, Structural, morphological, chemical and tribological investigations of sputter deposited MoS_x/metal multilayer coatings. *Surf. Coat. Technol.* **108–109**, 340–344 (1998)
- T. Spalvins, A review of recent advances in solid film lubrication. *J. Vac. Sci. Technol. A* **5**, 212–219 (1987)
- T. Spalvins, Lubrication with sputtered MoS₂ films: principles, operations, limitations. *J. Mater. Eng. Perfor.* **1**, 347–351 (1992)
- G. Xu, Z. Zhou, J. Liu, X. Ma, An investigation of fretting behavior of ion-plated TiN, magnetron-sputtered MoS₂ and their composite coatings. *Wear* **225–229**, 46–52 (1999)
- X. Zhang, R.G. Vitchev, W. Lauwerens, L. Stals, J. He, J.P. Celis, Effect of crystallographic orientation on fretting wear behavior of MoS_x coatings in dry and humid air. *Thin Solid Films* **396**, 69–77 (2001)

Motion of Geometry

- [Kinematics of Rolling Element Bearings](#)

Motor Oil

- [Engine Lubricants](#)

Motor Oil Test Equipment

- [Engine Oil Test Equipment](#)

Moving Mechanical Assemblies for Space Applications

- [Tribotechnology in Space Lubrication](#)

MPP – Modulated Pulsed Power (Trademarked)

- [PVD: Cathodic Arc and High Power Impulse Magnetron Sputtering \(HIPIMS\)](#)

MU, Measurement Uncertainty

- [Accuracy of Surface Topography Characterization Tools](#)

Multi-Grid Method

PEIRAN YANG

School of Mechanical Engineering, Qingdao Technological University, Qingdao, People's Republic of China

Definition

The multi-grid method is a kind of numerical method for solving partial differential equations or equation systems.

The concept of the multi-grid method is based on a certain understanding of the nature of slowly converging errors in iterative schemes. It is generally known that errors that have high frequencies are fast to converge on a domain with fine grids, however, errors that have low frequencies are fast to converge on the domain with coarse grids. Therefore, if a problem is solved iteratively on the same domain with different grids (i.e., with a multi-grid), all error components should be able to be quickly eliminated (Brandt 1977). Ever since the development of the first multilevel solver for elliptical partial differential equations, multilevel techniques have been developed for the solution of a wide variety of scientific problems. In the regime of tribology, the multi-grid method has been used successfully for the numerical solution of many kinds of elastohydrodynamic lubrication (EHL) problems. The first multi-grid EHL solver was developed by Lubrecht et al. (1986), and then improved by Venner et al. (1990, 1991). Although the multi-grid method has been used by numerous workers for different EHL problems, the principle and main techniques of their solvers are basically followed those developed by Lubrecht and Venner.

Scientific Fundamentals

The simplest EHL problem, i.e., the steady state, isothermal, Newtonian fluid lubricated line contact EHL problem, is taken as an example in this entry to describe the scientific fundamentals of the multi-grid method. Define the dimensionless variables as $P = p/p_H$, $H = hR/a^2$, $X = x/a$, $\bar{\eta} = \eta/\eta_0$, and $\bar{\rho} = \rho/\rho_0$, where p is the fluid pressure, h the film thickness, x the coordinate in the direction of entrainment, R the effective radius of curvature of rollers, p_H and a are the maximum Hertzian contact pressure and the half width of the Hertzian contact region, η and ρ the viscosity and density of the lubricant, and η_0 and ρ_0 the ambient viscosity and density. The dimensionless equations are as follows.

The Reynolds equation reads

$$\frac{d}{dX} \left(\varepsilon \frac{dP}{dX} \right) - \frac{d(\bar{\rho}H)}{dX} = 0, \quad (1)$$

in which

$$\varepsilon = \frac{\bar{\rho}H^3}{\bar{\eta}\lambda}, \quad \lambda = \frac{3}{4}\pi^2 U/W^2, \quad U = \frac{\eta_0 u}{E'R}, \quad W = \frac{w}{E'R},$$

where u is the entrainment velocity, E' the effective elastic modulus, and w the applied load of per unit length. If the inlet and outlet locations of the domain are denoted by X_{in}

and X_{out} , respectively, the boundary conditions of (1) can be expressed as $P = 0$ at $X = X_{in}$ and X_{out} , and $P \geq 0$ in the entire domain.

The film thickness equation reads

$$H(X) = H_0 + \frac{X^2}{2} - \frac{1}{\pi} \int_{X_{in}}^{X_{out}} P(X') \ln|X - X'| dX'. \quad (2)$$

The load balance equation reads

$$\int_{X_{in}}^{X_{out}} P dX = \frac{\pi}{2}. \quad (3)$$

In this entry the pressure-viscosity relation proposed by Roelands (1966) and the pressure-density relation proposed by Dowson and Higginson (1966) are adopted. The dimensionless expressions of these relations are

$$\bar{\eta} = \exp \left\{ (\ln \eta_0 + 9.67) \left[(1 + 5.1 \times 10^{-9} p_H P)^Z - 1 \right] \right\}, \quad (4)$$

and

$$\bar{\rho} = 1 + \frac{0.6 \times 10^{-9} p_H P}{1 + 1.7 \times 10^{-9} p_H P}. \quad (5)$$

In (4), Z is a dimensionless constant, its relation with the Barus' viscosity-pressure coefficient α is

$$Z = \alpha / [5.1 \times 10^{-9} (\ln \eta_0 + 9.67)].$$

Among the above five equations, only (1) and (3) are independent equations; the corresponding unknowns are pressure P and constant H_0 . Therefore, only these two equations need to be solved with the multi-grid method.

When the multi-grid method is used in the above model, several grids with different mesh sizes need to be arranged on the domain $\{X_{in}, X_{out}\}$. These grids can be considered as a single structure with several levels, of which the level with the least nodes is level 1, and that with the most nodes is level m . Equal-distance mesh is adopted on each level. If superscript k is used to indicate the level number, and the mesh size is denoted by Δ , then usually $\Delta^{k+1} = \Delta^k/2$ ($k = 1, 2, \dots, m-1$). In such a way, if the nodes on level k are $X_{in} = X_0^k, X_1^k, X_2^k, \dots, X_{n^k}^k = X_{out}$, then usually $n^{k+1} = 2n^k$ ($k = 1, 2, \dots, m-1$).

On level k , the Reynolds equation needs to be discretized into a set of algebraic equations with respect to the nodal pressures. This equation set can be written in a vector form as

$$L^k \mathbf{P}^k = \mathbf{f}^k, \quad (6)$$

where L^k is the discretized differential operator, \mathbf{P}^k the vector of nodal pressures, and \mathbf{f}^k the vector of the

right-hand terms, which should be considered as a known vector when \mathbf{P}^k is relaxed on level k . According to (1), on the finest level, i.e., when $k = m$, \mathbf{f}^k is a zero vector. When $k < m$, however, the right-hand vector should be determined with the following relation when and only when the relaxation on level k has just been finished and the process will immediately go down to level $k - 1$:

$$\mathbf{f}^{k-1} = L^{k-1}(\mathbf{I}_k^{k-1} \tilde{\mathbf{P}}^k) + \mathbf{I}_k^{k-1}(\mathbf{f}^k - L^k \tilde{\mathbf{P}}^k), \quad (7)$$

where $\tilde{\mathbf{P}}^k$ is the approximate solution just obtained on level k , and \mathbf{I}_k^{k-1} is a linear inter-level operator restricting results from level k to level $k - 1$. If \mathbf{u} is the vector and needs to be restricted, and subscripts $i = 2I$, the operator \mathbf{I}_k^{k-1} can be expressed as

$$\begin{cases} u_I^{k-1} = u_i^k & (I = 0 \text{ and } n^{k-1}) \\ u_I^{k-1} = \frac{1}{4}u_{i-1}^k + \frac{1}{2}u_i^k + \frac{1}{4}u_{i+1}^k & (I = 1, 2, \dots, n^{k-1} - 1) \end{cases} \quad (8)$$

Another interlevel operator is the linear interpolation operator, \mathbf{I}_k^{k+1} , which transfers results from level k to level $k + 1$. Still subscripts $i = 2I$, and the operator \mathbf{I}_k^{k+1} can be expressed as

$$\begin{cases} u_i^{k+1} = u_i^k & (I = 0, 1, \dots, n^k) \\ u_{i+1}^{k+1} = \frac{1}{2}u_i^k + \frac{1}{2}u_{i+1}^k & (I = 0, 1, \dots, n^k - 1) \end{cases}. \quad (9)$$

When a central finite difference scheme and a one-order backward finite difference scheme are applied to the first and second terms of (1), respectively, the detailed expression of (6) at node i on level k reads

$$\begin{aligned} & \frac{1}{(\Delta^k)^2} \left[\varepsilon_{i-1/2}^k P_{i-1}^k - (\varepsilon_{i-1/2}^k + \varepsilon_{i+1/2}^k) P_i^k + \varepsilon_{i+1/2}^k P_{i+1}^k \right] \\ & - \frac{1}{\Delta^k} (\bar{\rho}_i^k H_i^k - \bar{\rho}_{i-1}^k H_{i-1}^k) = f_i^k \quad (i = 1, 2, \dots, n^k - 1), \end{aligned} \quad (10)$$

where f_i^k is evaluated after the relaxation on the upper level, and $\varepsilon_{i-1/2}^k$ and $\varepsilon_{i+1/2}^k$ are given by

$$\varepsilon_{i-1/2}^k = \frac{1}{2}(\varepsilon_{i-1}^k + \varepsilon_i^k), \quad \varepsilon_{i+1/2}^k = \frac{1}{2}(\varepsilon_i^k + \varepsilon_{i+1}^k). \quad (11)$$

Since the film thickness equation is not independent, its discretized expression on level k can be derived directly from (2):

$$H_i^k = H_0 + \frac{(X_i^k)^2}{2} - \frac{1}{\pi} \sum_{j=0}^{n^k} K_{i,j}^k P_j^k \quad (i = 0, 1, \dots, n^k), \quad (12)$$

where

$$\begin{aligned} K_{i,j}^k = & \left(i - j + \frac{1}{2} \right) \Delta^k \left[\ln \left(\left| i - j + \frac{1}{2} \right| \Delta^k \right) - 1 \right] \\ & - \left(i - j - \frac{1}{2} \right) \Delta^k \left[\ln \left(\left| i - j - \frac{1}{2} \right| \Delta^k \right) - 1 \right]. \end{aligned} \quad (13)$$

Similarly, the discretized expressions of viscosity and density on level k can be written directly from (4) and (5) as

$$\begin{aligned} \bar{\eta}_i^k = & \exp \left\{ (\ln \eta_0 + 9.67) \left[(1 + 5.1 \times 10^{-9} p_H P_i^k)^Z - 1 \right] \right\} \\ & (i = 0, 1, \dots, n^k); \end{aligned} \quad (14)$$

$$\bar{\rho}_i^k = 1 + \frac{0.6 \times 10^{-9} p_H P_i^k}{1 + 1.7 \times 10^{-9} p_H P_i^k} \quad (i = 0, 1, \dots, n^k). \quad (15)$$

A local Newton-Raphson linearization combined with a Gauss-Seidel under-relaxation scheme is employed to solve the pressure. Once the H_i^k , $\bar{\eta}_i^k$, and $\bar{\rho}_i^k$ at each node have been calculated from (12), (14), and (15) with the given pressure vector \mathbf{P}^k , the new pressure vector $\tilde{\mathbf{P}}^k$ can be obtained from the following formulations:

$$\begin{cases} \tilde{\mathbf{P}}_i^k = \bar{\mathbf{P}}_i^k + \omega_1 \delta_i^k \\ \tilde{\mathbf{P}}_i^k = 0 & \text{if } \tilde{\mathbf{P}}_i^k < 0 \end{cases} \quad (i = 1, 2, \dots, n^k - 1), \quad (16)$$

where ω_1 is the under-relaxation factor, and

$$\delta_i^k = r_i^k / \left(\frac{\partial L_i^k}{\partial P_i^k} \right), \quad (17)$$

in which, the residual of the Reynolds equation is calculated by

$$\begin{aligned} r_i^k = & f_i^k - \left[\varepsilon_{i-1/2}^k \tilde{\mathbf{P}}_{i-1}^k - (\varepsilon_{i-1/2}^k + \varepsilon_{i+1/2}^k) \bar{\mathbf{P}}_i^k + \varepsilon_{i+1/2}^k \tilde{\mathbf{P}}_{i+1}^k \right] / \\ & (\Delta^k)^2 + (\bar{\rho}_i^k H_i^k - \bar{\rho}_{i-1}^k H_{i-1}^k) / \Delta^k, \end{aligned} \quad (18)$$

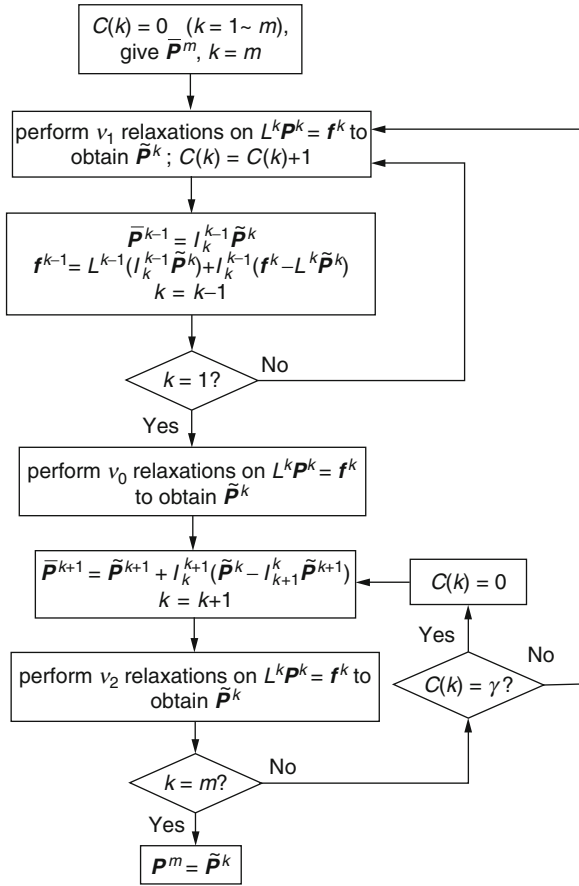
and the local Jacobi coefficient in (17) is calculated by

$$\begin{aligned} \frac{\partial L_i^k}{\partial P_i^k} = & - \left(\varepsilon_{i-1/2}^k + \varepsilon_{i+1/2}^k \right) / (\Delta^k)^2 \\ & + \frac{1}{\pi} \left(\bar{\rho}_i^k K_{i,i}^k - \bar{\rho}_{i-1}^k K_{i-1,i}^k \right) / \Delta^k. \end{aligned} \quad (19)$$

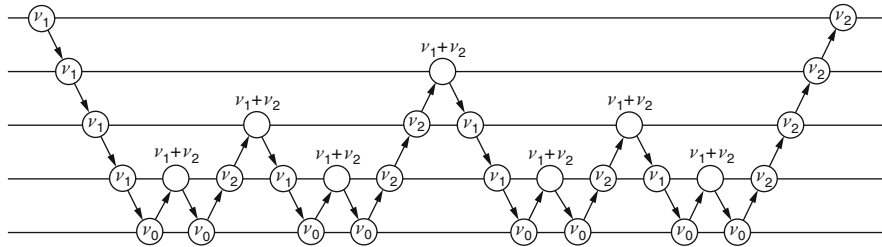
The nodal pressures are relaxed on different levels according to some chosen rules. The so-called W cycle is very suitable for EHL problems. The flowchart of a W cycle is shown in Fig. 1 (the process is a W cycle if $\gamma = 2$, while it is a V cycle if $\gamma = 1$). Furthermore, details of the W cycle for $m = 5$ are shown in Fig. 2. The numbers of relaxations are chosen based on experience, and usually $v_0 = 24$, $v_1 = 2$, and $v_2 = 1$ is a good set of choice.

The load balance equation, i.e., (3), is an independent equation controlling the unknown constant H_0 in (2). The discretized form of this equation on level k reads

$$0.5\Delta^k \sum_{i=0}^{n^k-1} (P_i^k + P_{i+1}^k) = g^k \quad (k = 1, 2, \dots, m), \quad (20)$$



Multi-Grid Method, Fig. 1 Flowchart of a W cycle for pressure relaxation (if $\gamma = 2$)



Multi-Grid Method, Fig. 2 The W cycle for $m = 5$

where g^k is the right-hand term. It is equal to $\pi/2$ when $k = m$, otherwise it is determined together with the right-hand terms of the Reynolds equation as

$$g^{k-1} = 0.5\Delta^{k-1} \sum_{i=0}^{n^{k-1}-1} \left[\left(I_k^{k-1} \tilde{P}^k \right)_i + \left(I_k^{k-1} \tilde{P}^k \right)_{i+1} \right] + g^k - 0.5\Delta^k \sum_{i=0}^{n^k-1} \left(\tilde{P}_i^k + \tilde{P}_{i+1}^k \right). \quad (21)$$

Constant H_0 can be adjusted on the lowest level (i.e., level 1) with the following scheme:

$$H_0 = \tilde{H}_0 + \omega_2 \left[0.5\Delta^1 \sum_{i=0}^{n^1-1} \left(\tilde{P}_i^1 + \tilde{P}_{i+1}^1 \right) - g^1 \right], \quad (22)$$

where ω_2 is a positive constant, which can be considered as another relaxation factor of the numerical procedure. The adjustment of H_0 should be performed after several relaxations for pressure. For example, if $v_0 = 24$, adjustment can be performed after the 5th, 10th, 15th, and 20th pressure relaxations.

The Hertzian contact pressure is often initialized on the finest level as \tilde{P}^m , and, a guessed value of H_0 is required to commence the first W cycle. When ω_1 and ω_2 are chosen properly, convergent solution can usually be obtained after 2 ~ 5 W cycles if the criterion of convergence is given by

$$\sum_{i=0}^{n^m} |\tilde{P}_i^m - \bar{P}_i^m| / \sum_{i=0}^{n^m} \tilde{P}_i^m < 0.001 \quad (23)$$

The numerical procedure described above basically follows that of Venner et al. (1991). There are two main differences between the present procedure and that of Venner et al. (1991): (1) in the present procedure, the discretized film thickness equation does not have a right-hand term; and (2) in the present procedure the dipole relaxation for pressure is not employed.

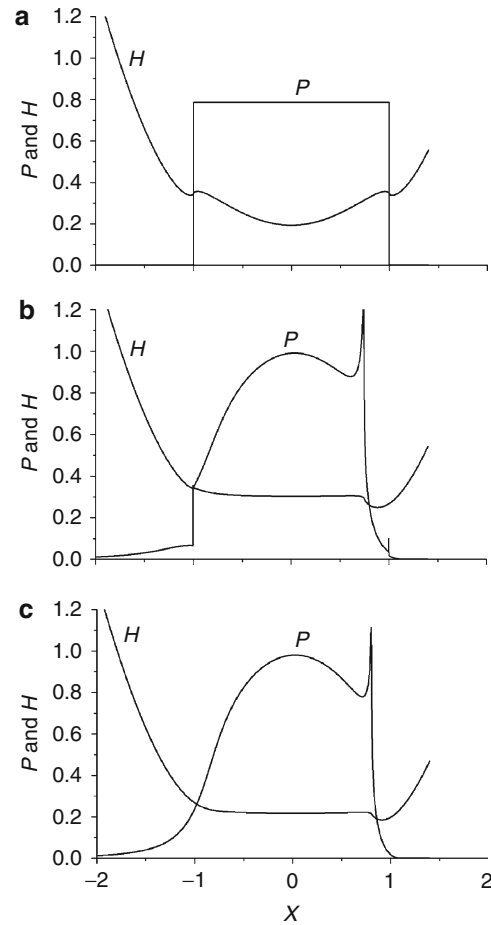
The multi-grid method is very fast. Running the code of the numerical solver described in this entry on

a personal computer with a central processing unit of Pentium Dual Core E2200, the time (not including the time for edit and link) consumed for a solution (with 3 W cycles) was only 13 s when $m = 10$ ($n^m = 15,360$); and was less than 1 s when m was less than 7 ($n^m < 1,920$). When the multilevel multi-integration technique was incorporated in the solver, the computing time can be even shorter (around 5 s when $m = 10$).

The multi-grid method has very good convergence property. With Hertzian contact pressure as \bar{P}^m , after the first W cycle, excluding the height of the pressure spike, the approximate solution is already very close to the convergent solution, that is, the difference cannot be distinguished in a figure. If the operating condition is not too severe, a convergent solution can even be achieved if \bar{P}^m is initialed by a step function. As an example, Fig. 3 shows the profiles of P and H at (a) the beginning of the first W cycle, (b) the end of the first W cycle, and (c) the end of the fourth W cycle (i.e., the convergent solution) for a typical line contact EHL case with $U = 1 \times 10^{-11}$, $G = 5,000$, and $W = 4 \times 10^{-5}$. It can be seen that the curves of P and H at the end of the first W cycle have already obtained the main EHL features, such as the pressure spike, the parallel gap, and the exit restriction of the film thickness.

The multi-grid method is very robust. If the relaxation factors ω_1 and ω_2 are chosen properly, and the domains are large enough to ensure the fully flooded conditions, convergent solutions can be obtained in very wide ranges of operating parameters. For $G = 5,000$ ($\alpha = 2.2 \times 10^{-8} \text{ Pa}^{-1}$), the suggested values of ω_1 and ω_2 together with the values of X_{in} and X_{out} are listed in Table 1 for 20 cases covering the major part of the U - W plane. Note that for cases 1 and 2, the maximum Hertzian pressure is as high as $p_H = 9.07 \text{ GPa}$. When the Hertzian pressure is taken as the testing pressure, and the testing value of H_0 is determined properly (the central film thickness given by an empirical formulas minus the central deformation produced by the initialed pressure), convergent solutions can be achieved with no more than 5 (mostly only 3) W cycles for all cases shown in Table 1 with $m = 6$ ($n^m = 960$).

In order to show the influence of the level number m on the accuracy of results, numerical solutions were obtained with $m = 3$ to $m = 10$ for the case with $U = 1 \times 10^{-11}$, $G = 5,000$, and $W = 4 \times 10^{-5}$. The predicted values of the dimensionless central film thickness, H_{cen} , dimensionless minimum film thickness, H_{min} , and the height of the dimensionless pressure spike, P_{spike} , are listed in Table 2. Furthermore, the profiles of the spikes for $m = 4$ –10 are plotted in Fig. 4. It can be seen that, both H_{cen} and H_{min} decrease as m increases, indicating that when the



Multi-Grid Method, Fig. 3 P and H versus X at: (a) the beginning of the first W cycle; (b) the end of the first W cycle; and (c) the end of the fourth W cycle when P is initialed as a step function for a typical EHL case with $U = 1 \times 10^{-11}$, $G = 5,000$, and $W = 4 \times 10^{-5}$

multi-grid method is applied to solve EHL problems, both the iterative convergence and the mesh convergence (Venner 2005) should be checked. From Table 2 and Fig. 4, it is also known that very dense meshes are required to predict the accurate height and shape of the pressure spike.

Key Applications

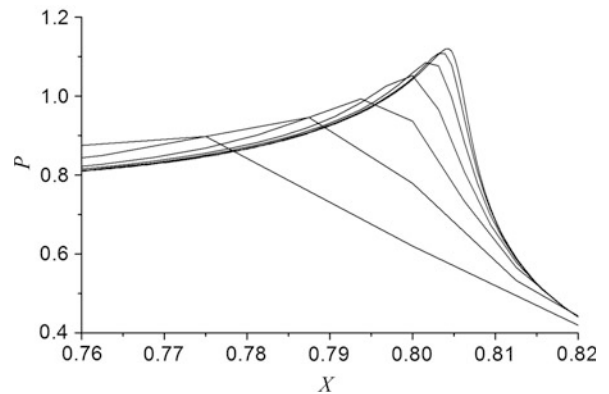
The multi-grid method has been applied to solve all kinds of EHL problems, especially for point contacts. For example, Yang et al. (2001) have solved the point contact thermal EHL problem to explain the dimple phenomenon; Hu and Zhu (2000) have studied the mixed lubrication in point contacts for both smooth and rough surfaces; Venner et al. (2008) have investigated the oil starvation

Multi-Grid Method, Table 1 Suggested Values of ω_1 , ω_2 , X_{in} and X_{out} ($G = 5,000$)

Case	U	W	ω_1	ω_2	X_{in}	X_{out}
1	1×10^{-11}	1×10^{-2}	0.3	0.02	-4.6	1.4
2	1×10^{-10}	1×10^{-2}	0.3	0.02	-4.6	1.4
3	1×10^{-13}	1×10^{-3}	0.3	0.02	-4.6	1.4
4	1×10^{-12}	1×10^{-3}	0.3	0.02	-4.6	1.4
5	1×10^{-11}	1×10^{-3}	0.3	0.02	-4.6	1.4
6	1×10^{-10}	1×10^{-3}	0.3	0.03	-4.6	1.4
7	1×10^{-13}	1×10^{-4}	0.3	0.02	-4.6	1.4
8	1×10^{-12}	1×10^{-4}	0.3	0.02	-4.6	1.4
9	1×10^{-11}	1×10^{-4}	0.3	0.04	-4.6	1.4
10	1×10^{-10}	1×10^{-4}	0.2	0.4	-9.2	2.8
11	1×10^{-13}	1×10^{-5}	0.3	0.02	-4.6	1.4
12	1×10^{-12}	1×10^{-5}	0.3	0.03	-4.6	1.4
13	1×10^{-11}	1×10^{-5}	0.5	0.05	-4.6	1.4
14	1×10^{-10}	1×10^{-5}	0.6	0.2	-9.2	2.8
15	1×10^{-9}	1×10^{-5}	1.0	2.5	-27.6	8.4
16	1×10^{-13}	1×10^{-6}	0.8	0.4	-9.2	2.8
17	1×10^{-12}	1×10^{-6}	0.8	0.5	-9.2	2.8
18	1×10^{-11}	1×10^{-6}	0.8	0.5	-18.4	5.6
19	1×10^{-10}	1×10^{-6}	1.0	5.0	-36.8	11.2
20	1×10^{-9}	1×10^{-6}	1.0	20.0	-73.6	22.4

Multi-Grid Method, Table 2 Effects of m (and n^m) on H_{cen} , H_{min} , and P_{spike} for $U = 1 \times 10^{-11}$, $G = 5,000$, and $W = 4 \times 10^{-5}$

m	n^m	H_{cen}	H_{min}	P_{spike}
3	120	0.229	0.191	0.864
4	240	0.224	0.189	0.898
5	480	0.221	0.187	0.947
6	960	0.219	0.186	0.994
7	1,920	0.219	0.185	1.052
8	3,840	0.218	0.185	1.085
9	7,680	0.218	0.185	1.109
10	15,360	0.218	0.185	1.120

**Multi-Grid Method, Fig. 4** Pressure spikes predicted with $m = 4$ to $m = 10$ for $U = 1 \times 10^{-11}$, $G = 5,000$, and $W = 4 \times 10^{-5}$

phenomenon in point EHL contacts; and Liu et al. (2005) have developed a solver for non-Newtonian thermal EHL in point contacts. The multi-grid method is so powerful that it can even be used to solve very complicated EHL problem, for example, Yang et al. (2008) have just solved a point EHL contact problem, in which transient and thermal effects, non-Newtonian flow, and surface

roughness have been taken into account simultaneously. Multi-grid method has also found application in spherical coordinate system for artificially human joints (Gao et al. 2007).

It should be pointed out that the multilevel multi-integration (MLMI) algorithm or the fast Fourier

transform (FFT) technique has to be incorporated in the multi-grid solvers if the multi-grid method is used to solve EHL problems in point contacts, otherwise, the time of computation will be unbearably long.

References

- A. Brandt, Multi-level adaptive solutions to boundary-value problems. *Math. Comp.* **31**, 333–390 (1977)
- D. Dowson, G.R. Higginson, *Elastohydrodynamic Lubrication* (Pergamon, Oxford, 1966)
- L. Gao et al., Comparison of numerical methods for elastohydrodynamic lubrication analysis of metal-on-metal hip implants: multi-grid versus Newton-Raphson. *Proc. IMechE Part J J. Eng. Tribol.* **221**, 133–140 (2007)
- Y.Z. Hu, D. Zhu, A full numerical solution to the mixed lubrication in point contacts. *ASME J. Tribol.* **122**, 1–9 (2000)
- X. Liu et al., Non-Newtonian thermal analyses of point EHL contacts using the Eyring model. *ASME J. Tribol.* **127**, 70–81 (2005)
- A.A. Lubrecht et al., Multigrid, an alternative method for calculating film thickness and pressure profiles in elastohydrodynamically lubricated line contacts. *ASME J. Tribol.* **108**, 551–556 (1986)
- C.J.A. Roelands, Correlation Aspects of Viscosity-Temperature-Pressure Relationship of Lubricating Oils, Ph.D. thesis, Delft University of Technology, Delft, 1966
- C.H. Venner, EHL film thickness computations at low speeds: risk of artificial trends as a results of poor accuracy and implications for mix lubrication modeling. *Proc. IMechE Part J: J. Eng. Tribol.* **219**, 285–290 (2005)
- C.H. Venner et al., Advanced multilevel solution of the EHL line contact problem. *ASME J. Tribol.* **112**, 426–431 (1990)
- C.H. Venner, Multilevel Solution of the EHL Line and Point Contact Problems, Ph. D. thesis, The University of Twente, Enschede, 1991
- C.H. Venner et al., Film thickness modulations in starved elastohydrodynamically lubricated contacts induced by time-varying lubricant supply. *ASME J. Tribol.* **130**, 041501 (2008)
- P. Yang et al., Formation of steady dimples in point TEHL contacts. *ASME J. Tribol.* **123**, 42–49 (2001)
- P. Yang et al., Influence of two-sided surface waviness on the EHL behavior of rolling/sliding point contacts under thermal and non-Newtonian conditions. *ASME J. Tribol.* **130**, 041502 (2008)

Multi-Level Multi-Integration for Deformation

PEIRAN YANG

School of Mechanical Engineering,
Qingdao Technological University, Qingdao,
People's Republic of China

Definition

Multi-level multi-integration (MLMI) is an algorithm for the fast solution of a kind of integral equations (Brandt and Lubrecht 1990). In the regime of tribology, especially

in the theoretical studies of elastohydrodynamic lubrication (EHL), MLMI is mainly used to calculate the elastic deformation of contacting surfaces (Venner et al. 1991). In dimensionless form, the elastic deformation E over a rectangular domain Ω for a point contact EHL problem needs to be solved from the following integral equation:

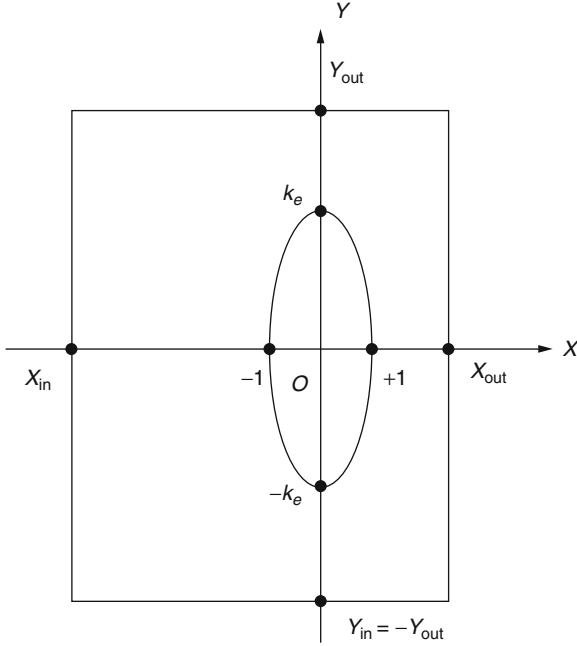
$$E(X, Y) = A \iint_{\Omega} \frac{P(X', Y')}{\sqrt{(X - X')^2 + (Y - Y')^2}} dX' dY', \quad (1)$$

where A is a constant and P the dimensionless pressure. To calculate $E(X, Y)$ at a certain location requires an integration over the entire domain. The discrete analogue of multi-integration is the multiplication of the vector of nodal pressures by a dense matrix coming from the integration kernel, which has a singular point at $(X = X', Y = Y')$. If (1) is discretized in the X and Y directions with uniform mesh sizes ΔX and ΔY , the number of numerical operations for the deformation at each node is proportional to the number of all nodes, n . Over the entire domain, therefore, the total number of numerical operations is proportional to n^2 if conventional grid (with only one level) is used. When multi-level technique is used, however, the integration is performed on a selected grid level with coarse mesh, and on the levels with finer meshes the integration is obtained through level-by-level interpolation and the necessary correction because of the singularity of the integration kernel. In such a way, the complexity of the algorithm can be reduced from $O(n^2)$ to $O(n \ln n)$, therefore, without losing accuracy, the efficiency of MLMI is much higher than that of the conventional algorithm.

Scientific Fundamentals

Only the MLMI for two-dimensional problems will be described in this entry. The dimensionless coordinate system, the domain for the solution, and a contact ellipse are shown in Fig. 1. In such a domain, (1) will be solved numerically with the MLMI algorithm. Note in (1) and Fig. 1, $E = d_s R_x / a^2$, $P = p / p_H$, $X = x / a$, $Y = y / a$, and $k_e = b / a$, where a and b are the minor radius and major radius of the Hertzian contact ellipse, respectively, d_s is the sum of the elastic deformations of two solid surfaces, R_x the effective radius of curvature of contact bodies measured on the plane passing the x -axis, and p_H the maximum Hertzian contact pressure at a constant load supported by the contact.

Several levels of grids with different mesh sizes need to be arranged on the domain. These grids can be considered as a single structure, of which the level with the least nodes is level 1, and that with the most nodes is level m .



Multi-Level Multi-Integration for Deformation, Fig. 1
The domain of (1) along with a Hertzian contact ellipse

If superscript k is used to indicate the level number, and the uniform mesh sizes are denoted by ΔX and ΔY in the X and Y directions, respectively, the relations of the mesh sizes on level k and level $k+1$ are: $\Delta X^{k+1} = \Delta X^k/2$, and $\Delta Y^{k+1} = \Delta Y^k/2$ ($k = 1, 2, \dots, m-1$). In such a way, if on level k the nodal locations in the X direction are $X_0^k, X_1^k, X_2^k, \dots, X_{n_X^k}^k$, and in the Y direction are $Y_0^k, Y_1^k, Y_2^k, \dots, Y_{n_Y^k}^k$, then usually $n_X^{k+1} = 2n_X^k$, and $n_Y^{k+1} = 2n_Y^k$ ($k = 1, 2, \dots, m-1$).

Neglecting superscript k , the nodal deformation on level k can be expressed as

$$E_{i,j} = A \sum_{k=0}^{n_X} \sum_{l=0}^{n_Y} K_{ij,kl} P_{k,l} \quad (i = 0 \sim n_X, j = 0 \sim n_Y), \quad (2)$$

where $(K_{ij,kl})$ is a four-dimensional array, its elements are

$$K_{ij,kl} = E_1 + E_2 + E_3 + E_4, \quad (3)$$

where

$$\begin{cases} E_1 = \Psi_1 \ln \frac{\Gamma_1 + \sqrt{\Gamma_1^2 + \Psi_1^2}}{\Gamma_2 + \sqrt{\Gamma_2^2 + \Psi_1^2}}, & E_2 = \Psi_2 \ln \frac{\Gamma_2 + \sqrt{\Gamma_2^2 + \Psi_2^2}}{\Gamma_1 + \sqrt{\Gamma_1^2 + \Psi_2^2}}, \\ E_3 = \Gamma_1 \ln \frac{\Psi_1 + \sqrt{\Gamma_1^2 + \Psi_1^2}}{\Psi_2 + \sqrt{\Gamma_1^2 + \Psi_2^2}}, & E_4 = \Gamma_2 \ln \frac{\Psi_2 + \sqrt{\Gamma_2^2 + \Psi_2^2}}{\Psi_1 + \sqrt{\Gamma_2^2 + \Psi_1^2}}, \\ \Gamma_1 = (k - i + \frac{1}{2})\Delta X, & \Gamma_2 = (k - i - \frac{1}{2})\Delta X, \\ \Psi_1 = (l - j + \frac{1}{2})\Delta Y, & \Psi_2 = (l - j - \frac{1}{2})\Delta Y. \end{cases} \quad (4)$$

It can be noticed that array $(K_{ij,kl})$ has numerous identical elements, therefore, it can be reduced into a square matrix $[D_{|i-k|,|j-l|}]$, in which

$$D_{|i-k|,|j-l|} = K_{ij,kl} \quad (|i-k| = 0 \sim n_X, |j-l| = 0 \sim n_Y). \quad (5)$$

The actual ranges of $|i-k|$ and $|j-l|$ are a little larger than those described in (5), because some virtual nodes are required by the following high-order operators of interpolation and restriction for transferring the nodal values of vector \mathbf{u} :

$$\mathbf{u}^{k+1} = \Pi_k^{k+1} \mathbf{u}^k, \quad (6)$$

and

$$\mathbf{u}^k = \Pi_{k+1}^k \mathbf{u}^{k+1}, \quad (7)$$

where, Π_k^{k+1} is the operator of interpolation, while Π_{k+1}^k is the operator of restriction. If the subscripts are expressed by I and J on level k , while by i and j on level $k+1$, detailed operations of Π_k^{k+1} are as follows:

$$u_{i,j}^{k+1} = \begin{cases} u_{I,J}^k & i = 2I, j = 2J, \\ \frac{1}{16} \left(-u_{I-1,J}^k + 9u_{I,J}^k + 9u_{I+1,J}^k - u_{I+2,J}^k \right) & i = 2I+1, j = 2J, \\ \frac{1}{16} \left(-u_{I,J-1}^k + 9u_{I,J}^k + 9u_{I,J+1}^k - u_{I,J+2}^k \right) & i = 2I, j = 2J+1, \\ \frac{81}{256} \left(u_{I,J}^k + u_{I+1,J+1}^k + u_{I+1,J}^k + u_{I,J+1}^k \right) & \\ - \frac{9}{256} \left(u_{I-1,J}^k + u_{I+2,J}^k + u_{I-1,J+1}^k + u_{I+2,J+1}^k \right. & \\ \quad \left. + u_{I,J-1}^k + u_{I,J+2}^k + u_{I+1,J-1}^k + u_{I+1,J+2}^k \right) & \\ + \frac{1}{256} \left(u_{I-1,J-1}^k + u_{I+2,J+2}^k + u_{I-1,J+2}^k + u_{I+2,J-1}^k \right) & \\ i = 2I+1, j = 2J+1. & \end{cases} \quad (8)$$

The corresponding restricting operations of Π_{k+1}^k are

$$u_{I,J}^k = \frac{1}{4} \left[u_{i,j}^{k+1} + \frac{9}{16} \left(u_{i-1,j}^{k+1} + u_{i,j-1}^{k+1} + u_{i+1,j}^{k+1} + u_{i,j+1}^{k+1} \right) \right. \\ + \frac{81}{256} \left(u_{i-1,j-1}^{k+1} + u_{i-1,j+1}^{k+1} + u_{i+1,j-1}^{k+1} + u_{i+1,j+1}^{k+1} \right) \\ + \frac{1}{256} \left(u_{i-3,j-3}^{k+1} + u_{i-3,j+3}^{k+1} + u_{i+3,j-3}^{k+1} + u_{i+3,j+3}^{k+1} \right) \\ - \frac{9}{256} \left(u_{i-3,j-1}^{k+1} + u_{i-3,j+1}^{k+1} + u_{i-1,j-3}^{k+1} + u_{i-1,j+3}^{k+1} \right. \\ \left. + u_{i+1,j-3}^{k+1} + u_{i+1,j+3}^{k+1} + u_{i+3,j-1}^{k+1} + u_{i+3,j+1}^{k+1} \right) \\ \left. - \frac{1}{16} \left(u_{i-3,j}^{k+1} + u_{i,j-3}^{k+1} + u_{i,j+3}^{k+1} + u_{i+3,j}^{k+1} \right) \right] \\ i = 2I, j = 2J. \quad (9)$$

An important point of MLMI is the concept of two special levels, i.e., the aimed level denoted by k_a , and the integration level denoted by k_d .

The aimed level is the level on which the pressure happens to be relaxed in the multi-grid EHL solver; therefore, any level can be the aimed level because in the multi-grid procedure the pressure is relaxed alternatively on all levels according to a chosen rule such as the W cycle. The integration level is the level on which the nodal deformations are integrated over the entire domain. Usually, the integration level is a fixed coarse level and can be chosen as $k_d = 2$ if $m \leq 6$ or $k_d = 3$ if $m > 6$.

When $k_a \leq k_d$, the nodal deformations on the aimed level are calculated directly from (2) to (4).

Coarse Grid Integration When $k_a > k_d$ the nodal deformations on the integration level are calculated by

$$E_{I,J}^{k_d} = \sum_{K=0}^{n_X^{k_d}} \sum_{L=0}^{n_Y^{k_d}} \tilde{K}_{IJ,KL}^{k_d} \hat{P}_{K,L}^{k_d}. \quad (10)$$

The coefficients $\tilde{K}_{IJ,KL}^{k_d}$ are not calculated from (2) to (5) on level k_d but on level k_a , the obtained coefficients are then injected (i.e., taking the values from the coincident nodes) to the nodes on level k_d . Furthermore, values of $\hat{P}_{IJ,KL}^{k_d}$ come from the approximate solution of the pressure vector $\tilde{\mathbf{P}}^{k_a}$ on level k_a level-by-level down until level k_d with the following operation:

$$\mathbf{u}^k = 4 \times \Pi_{k+1}^k \mathbf{u}^{k+1}. \quad (11)$$

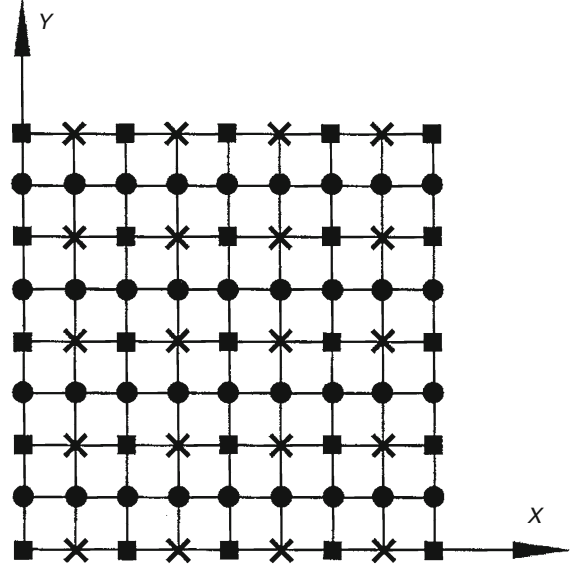
Fine Grid Correction The nodes on level $k_d + 1$ can be divided into three classes, see Fig. 2, where the nodes of different classes are marked with different symbols.

The first-class nodes are those marked with dark squares in Fig. 2. The locations of these nodes are coincident with the nodes on level k_d . If subscripts i and j are used on level $k_d + 1$ while subscripts I and J are used on level k_d for the first class nodes, $i = 2I$ and $j = 2J$. When the integration on level k_d has finished, the deformations of all first-class nodes (including the necessary virtual nodes outside the domain) should be corrected firstly as follows:

$$E_{i,j}^{k_d+1} = \tilde{E}_{i,j}^{k_d+1} + \sum_{k=i-S_X}^{i+S_X} \sum_{l=j-S_Y}^{j+S_Y} C1_{ij,kl}^{k_d+1} \hat{P}_{k,l}^{k_d+1}, \quad (12)$$

where $\tilde{E}_{i,j}^{k_d+1} = E_{I,J}^{k_d}$, see the formula for $i = 2I$ and $j = 2J$ in (8). $\hat{P}_{k,l}^{k_d+1}$ is the nodal pressure on level $k_d + 1$ if $k_d + 1 = k_a$, otherwise it comes from the upper level or levels in the way as obtaining $\hat{P}_{K,L}^{k_d}$ in (11).

In (12), S_X and S_Y are the numbers of nodes the correction needs to be performed around node



Multi-Level Multi-Integration for Deformation, Fig. 2 Three classes of nodes on level $k_d + 1$

$(X_i^{k_d+1}, Y_j^{k_d+1})$. Because the influence of the singularity of the integration kernel decays rapidly with the distance from the singular point, the values of S_X and S_Y are usually quite small. Numerical experiment shows that, the values S_X and S_Y given by the following equations can always maintain the accuracy of the integration:

$$\begin{cases} S_X = 5 + 2 \times \ln(n_X), \\ S_Y = 5 + 2 \times \ln(n_Y). \end{cases} \quad (13)$$

Note that superscript $k_d + 1$ has been ignored from (12) to (13) for S_X , S_Y and n_X , n_Y .

$C1_{ij,kl}^{k_d+1}$ in (12) is the element of the correction coefficient array $(C1_{ij,kl}^{k_d+1})^{k_a}$. This is a six-dimensional array because i, j, k, l , and k_a are variables, and k_d also needs to be considered as a variable because correction often needs to be performed on upper level or levels. There are many identical elements in array $(C1_{ij,kl}^{k_d+1})^{k_a}$, therefore, it can be reduced into a four-dimensional array ($\tilde{C}1$) by letting

$$C1_{ij,kl}^{k_d+1} = \tilde{C}1_{|i-k|,|j-l|}^{k_d+1}. \quad (14)$$

Denoting $|i - k| = i1$ and $|j - l| = j1$, the elements of ($\tilde{C}1$) can be derived as

$$\tilde{C}_1^{k_d+1} = \begin{cases} 0 & (\text{both } i1 \text{ and } j1 \text{ are even numbers}), \\ \tilde{D}_{i1,j1}^{k_d+1} + \frac{1}{16}\tilde{D}_{|i1-3|,j1}^{k_d+1} - \frac{9}{16}\tilde{D}_{|i1-1|,j1}^{k_d+1} - \frac{9}{16}\tilde{D}_{i1,j1+1}^{k_d+1} + \frac{1}{16}\tilde{D}_{i1+3,j1}^{k_d+1} & (i1 \text{ is odd number, } j1 \text{ is even number}), \\ \tilde{D}_{i1,j1}^{k_d+1} + \frac{1}{16}\tilde{D}_{i1,|j1-3|}^{k_d+1} - \frac{9}{16}\tilde{D}_{i1,|j1-1|}^{k_d+1} - \frac{9}{16}\tilde{D}_{i1,j1+1}^{k_d+1} + \frac{1}{16}\tilde{D}_{i1,j1+3}^{k_d+1} & (i1 \text{ is even number, } j1 \text{ is odd number}), \\ \tilde{D}_{i1,j1}^{k_d+1} - \frac{81}{256}\left(\tilde{D}_{i1+1,j1+1}^{k_d+1} + \tilde{D}_{|i1-1|,j1+1}^{k_d+1} + \tilde{D}_{i1+1,|j1-1|}^{k_d+1} + \tilde{D}_{|i1-1|,|j1-1|}^{k_d+1}\right) \\ + \frac{9}{256}\left(\tilde{D}_{i1+1,j1+3}^{k_d+1} + \tilde{D}_{|i1-1|,j1+3}^{k_d+1} + \tilde{D}_{i1+3,j1+1}^{k_d+1} + \tilde{D}_{|i1-1|,j1+1}^{k_d+1}\right) \\ + \tilde{D}_{i1+3,|j1-1|}^{k_d+1} + \tilde{D}_{|i1-3|,|j1-1|}^{k_d+1} + \tilde{D}_{i1+1,|j1-3|}^{k_d+1} + \tilde{D}_{|i1-1|,|j1-3|}^{k_d+1} \\ - \frac{1}{256}\left(\tilde{D}_{i1+3,j1+3}^{k_d+1} + \tilde{D}_{|i1-3|,j1+3}^{k_d+1} + \tilde{D}_{i1+3,|j1-3|}^{k_d+1} + \tilde{D}_{|i1-3|,|j1-3|}^{k_d+1}\right) & (\text{both } i1 \text{ and } j1 \text{ are odd numbers}). \end{cases} \quad (15)$$

It is necessary to point out that the values of \tilde{D} in (15) are calculated from (3) to (5) if $k_d + 1 = k_a$, otherwise they are injected from the values of D on level k_a .

When deformations of all first-class nodes have been corrected through (12), the deformations of these nodes can be considered as the same as those integrated directly on level $k_d + 1$, therefore, the corrected values can be called the ideal values.

The second class nodes are those marked with **x** in Fig. 2. The approximate values of the deformations at these nodes come from the interpolation of the ideal values of the first-class nodes in the X direction; see the formula for $i = 2I + 1$ and $j = 2J$ in (8). In detail, the approximate deformations are given by

$$\tilde{E}_{i,j}^{k_d+1} = -\frac{1}{16}E_{i-3,j}^{k_d+1} + \frac{9}{16}E_{i-1,j}^{k_d+1} + \frac{9}{16}E_{i+1,j}^{k_d+1} - \frac{1}{16}E_{i+3,j}^{k_d+1}. \quad (16)$$

In order to obtain the ideal values at these nodes, the deformations given by (16) also need to be corrected. The formula of correction reads

$$E_{i,j}^{k_d+1} = \tilde{E}_{i,j}^{k_d+1} + \sum_{k=i-S_X}^{i+S_X} \sum_{l=j-S_2}^{j+S_2} C_2^{k_d+1} \hat{P}_{k,l}^{k_d+1}, \quad (17)$$

where S_2 is the number of nodes for the correction in the direction perpendicular to the direction of interpolation. Numerical experiment shows that $S_2 = 6$ can always maintain the accuracy of MLMI. Again, the six-dimensional array $(C_2^{k_d+1})^{k_a}$ can be reduced into a four-dimensional array (\tilde{C}_2) by letting

$$C_2^{k_d+1} = \tilde{C}_2^{k_d+1}_{|i-k|,|j-l|}. \quad (18)$$

Denoting $|i - k| = i2$ and $|j - l| = j2$, the elements of (\tilde{C}_2) can be derived as

$$\tilde{C}_2^{k_d+1} = \tilde{D}_{i2,j2}^{k_d+1} + \frac{1}{16}\tilde{D}_{|i2-3|,j2}^{k_d+1} - \frac{9}{16}\tilde{D}_{|i2-1|,j2}^{k_d+1} - \frac{9}{16}\tilde{D}_{i2+1,j2}^{k_d+1} + \frac{1}{16}\tilde{D}_{i2+3,j2}^{k_d+1}. \quad (19)$$

The third class nodes are those marked with **•** in Fig. 2. The approximate values of the deformations of these nodes come from the interpolation of the ideal values of the first-class and second-class nodes in the Y direction; see the formula for $i = 2I$ and $j = 2J + 1$ in (8). In detail, the approximate deformations are given by

$$\tilde{E}_{i,j}^{k_d+1} = -\frac{1}{16}E_{i,j-3}^{k_d+1} + \frac{9}{16}E_{i,j-1}^{k_d+1} + \frac{9}{16}E_{i,j+1}^{k_d+1} - \frac{1}{16}E_{i,j+3}^{k_d+1}. \quad (20)$$

Similar to the treatment for the second class nodes, the formula of the correction for the third class nodes is

$$E_{i,j}^{k_d+1} = \tilde{E}_{i,j}^{k_d+1} + \sum_{k=i-S_2}^{i+S_2} \sum_{l=j-S_Y}^{j+S_Y} C_3^{k_d+1} \hat{P}_{k,l}^{k_d+1}. \quad (21)$$

And, let

$$C_3^{k_d+1} = \tilde{C}_3^{k_d+1}_{|i-k|,|j-l|}, \quad (22)$$

the six-dimensional array $(C_3^{k_d+1})^{k_a}$ can be reduced into a four-dimensional array (\tilde{C}_3) :

$$\tilde{C}_3^{k_d+1} = \tilde{D}_{i3,j3}^{k_d+1} + \frac{1}{16}\tilde{D}_{i3,|j3-3|}^{k_d+1} - \frac{9}{16}\tilde{D}_{i3,|j3-1|}^{k_d+1} - \frac{9}{16}\tilde{D}_{i3,j3+1}^{k_d+1} + \frac{1}{16}\tilde{D}_{i3,j3+3}^{k_d+1}. \quad (23)$$

where $i3 = |i - k|$, and $j3 = |j - l|$.

Correction on Upper Levels If $k_d + 1 \neq m$, level $k_d + 1$ can be considered as the level of integration, and the ideal values of nodal deformations on this level can be considered to be the same as the results of integration over the entire domain. Then, the nodal deformations on the upper level can be calculated in the same way as described above. This procedure should be continued until the aimed level has been reached.

Key Applications

The accuracy and efficiency of MLMI can be seen from a comparison between the results of running the multi-grid EHL programs with and without MLMI to solve a typical EHL problem. The isothermal steady-state point contact EHL problem with the dimensionless parameters (Hamrock and Dowson 1977)

$W = 2.578 \times 10^{-5}$ ($p_H = 1.5 \text{ GPa}$, $R_x = 0.02 \text{ m}$), $G = 4,972$ ($E' = 2.26 \times 10^{11} \text{ Pa}$), $U = 1 \times 10^{-11}$, and $k_e = 3$ was chosen for this purpose. Level numbers $m = 3$ to 7 were tested with $n_X^1 = n_Y^1 = 16$, therefore, the total nodes were $1,025 \times 1,025$ on the finest level for $m = 7$.

The time consumed on a personal computer with a central processing unit Pentium Dual Core E2200, the predicted central film thicknesses, h_{cen} , along with the numbers of W cycles performed for the convergent solutions are listed in Table 1. The numerical results were accepted as solutions when the relative error of pressures checked at the beginning and end of a W cycle and the error of load balance were less than 0.001, simultaneously.

The results in Table 1 indicate that MLMI has no influence on the accuracy of the solutions, although it can reduce the time of computation drastically, especially when very dense meshes are used. It appears from Table 1 that the ratio of the consumed time without and with MLMI has a simple rule versus the number of

levels. That is, the time ratios are 8^0 with $m = 3$, 8^1 with $m = 4$, 8^2 with $m = 5$, and 8^3 with $m = 6$. If this rule is true, it can be estimated that the time ratio with $m = 7$ will be 8^4 , therefore, without MLMI, the solution will spend about 13 days. Furthermore, the variation in h_{cen} versus m shown in Table 1 indicates that, in order to predict the EHL film thickness accurately, sometimes very dense grids, such as those with more than a million nodes, are required.

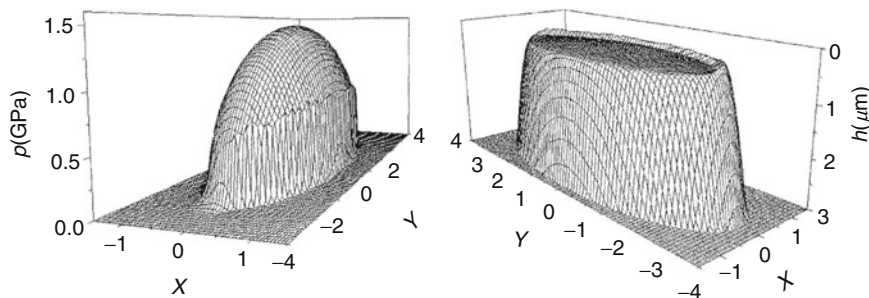
Three dimensional distributions of the pressure and film thickness obtained for the case in Table 1 with $m = 5$ are illustrated in Fig. 3 (some nodes have been skipped for clearance).

The algorithm of MLMI for deformation has been collaborated with the multi-grid method for pressure to solve almost all kinds of point contact EHL problems. For example, Qu et al. (2000) have solved the point contact thermal EHL problem to explain the dimple phenomenon discovered by Kaneta et al. (1992); Hu and Zhu (2000) have studied the mixed lubrication in point contacts for both smooth and rough surfaces; Venner et al. (2008) have investigated the oil starvation phenomenon in point EHL contacts; and Liu et al. (2005) have developed a solver for non-Newtonian thermal EHL in point contacts. Multi-grid solver cooperating with MLMI is so powerful that it can even be used to solve very complicated EHL problems, in which transient and thermal effects, non-Newtonian flow, and surface topography need to be taken into account simultaneously (Yang et al. 2008).

MLMI for deformation has also been applied to line contact EHL problems (Venner et al. 1990; Yang et al. 2001). However, with both the development of multi-level techniques and the progress of computer science, MLMI has gradually lost its importance for line contact EHL problems because the nodes used for line contact problems are too few to reflect the advantage of MLMI. This can be seen from a simple comparison: without MLMI, the running time of a multi-grid solver with

Multi-Level Multi-Integration for Deformation, Table 1
Levels of grids, performed W cycles, consumed time on a HP Pavilion computer, and the predicted central film thicknesses by the multi-grid solvers with and without MLMI, for an elliptical contact isothermal EHL problem with $W = 2.578 \times 10^{-5}$, $G = 4,972$, $U = 1 \times 10^{-11}$, and $k_e = 3$

m	W cycles	Time with MLMI	Time without MLMI	h_{cen} (μm) with MLMI	h_{cen} (μm) without MLMI
3	10	2 s	4 s	0.2898	0.2897
4	9	5 s	38 s	0.3588	0.3588
5	7	16 s	18 min	0.3600	0.3599
6	5	54 s	8 h	0.3397	0.3396
7	5	5 min	not test	0.3206	—



Multi-Level Multi-Integration for Deformation, Fig. 3 Pressure and film thickness solved with five levels of grids ($k_e = 3$, $U = 1 \times 10^{-11}$, $G = 4,972$, $p_H = 1.5 \text{ GPa}$, $R_x = 0.02 \text{ m}$)

960 meshes for a line contact EHL problem was 4 h on a scientific computer in 1985 (Lubrecht, et al. 1986), while it was less than 1 s (excluding the time for edit and link) on a personal computer in 2009, see the entry entitled ► [Multi-Grid Method](#) in this Encyclopedia.

References

- A. Brandt, A.A. Lubrecht, Multilevel matrix multiplication and fast solution of integral equations. *J. Comput. Phys.* **90**, 348–370 (1990)
- B.J. Hamrock, D. Dowson, Isothermal elastohydrodynamic lubrication of point contacts, part III – fully flooded results. *ASME J. Lubr. Technol.* **99**, 264–276 (1977)
- Y.Z. Hu, D. Zhu, A full numerical solution to the mixed lubrication in point contacts. *ASME J. Tribol.* **122**, 1–9 (2000)
- M. Kaneta et al., Effects of elastic moduli of contact surfaces in elastohydrodynamic lubrication. *ASME J. Tribol.* **114**, 75–80 (1992)
- X. Liu et al., Non-Newtonian thermal analyses of point EHL contacts using the Eyring model. *ASME J. Tribol.* **127**, 70–81 (2005)
- A.A. Lubrecht et al., Multigrid, an alternative method for calculating film thickness and pressure profiles in elastohydrodynamically lubricated line contacts. *ASME J. Tribol.* **108**, 551–556 (1986)
- S. Qu et al., Theoretical investigation on the dimple occurrence in the thermal EHL of simple sliding steel-glass circular contacts. *Tribol. Int.* **33**, 59–65 (2000)
- C.H. Venner et al., Advanced multilevel solution of the EHL line contact problem. *ASME J. Tribol.* **112**, 426–431 (1990)
- C.H. Venner, Multilevel solution of the EHL line and point contact problems, Ph.D. thesis, The University of Twente, Enschede, 1991
- C.H. Venner et al., Film thickness modulations in starved elastohydrodynamically lubricated contacts induced by time-varying lubricant supply. *ASME J. Tribol.* **130**, 041501 (2008)
- P. Yang et al., On the theory of thermal elastohydrodynamic lubrication at high slide-roll ratios – line contact solution. *ASME J. Tribol.* **123**, 36–41 (2001)
- P. Yang et al., Influence of two-sided surface waviness on the EHL behavior of rolling/sliding point contacts under thermal and non-Newtonian conditions. *ASME J. Tribol.* **130**, 041502 (2008)

Multiplex Coatings

M. DE JEFF T. HOSSON, YUTAO PEI
 Department of Applied Physics, Materials Innovation
 Institute and Zernike Institute for Advanced Materials,
 University of Groningen, Groningen, AG,
 The Netherlands

Definition

The tribological performance of nanocomposite diamond-like carbon (DLC) coatings and the breakdown of the Coulomb friction law in nanocomposite materials are discussed here. Physical arguments are provided to explain the dependence of friction on sliding velocity, in the

context of self-lubrication and influence of water vapor adsorption from the testing environment. From a structural design viewpoint, nanocomposite coatings exhibit an optimized combination of functionalities when a nanoparticle becomes approximately the same size as the separation.

Scientific Fundamentals

Introduction

Many laws in phenomenological materials physics are based on the concept that a constant driving force will lead to a response that is stable in time. In fact, the entire framework of solid state mechanics, founded by Isaac Newton himself, is based on this principle. In the field of materials science, this concept was believed to be true also. For example, a constant load will lead to a particular deformation response everywhere and at any time in the material; a constant frictional load will generate a constant sliding velocity for two surfaces in contact, and so on. Nevertheless, in recent times it has been demonstrated convincingly that in many situations such a concept does not apply. Typical examples being the jerky motion of dislocations when deforming a metal (De Hosson et al. 1983) or the jerky (stick-slip) motion of surfaces in relative motion (Kerssemakers and De Hosson 1998). Since the time of Leonardo da Vinci, who was arguably the first engineer to study friction in detail, stick-slip phenomena at interacting surfaces have become an important branch of modern materials science and engineering. Due to the complexity of friction processes combining individual physical events between sliding surfaces, it is still, however, a challenge to understand the precise mechanisms of stick-slip friction on a micrometer scale. Indeed friction between two surfaces in relative motion is a complex phenomenon that involves phonon dissipation, bond breaking and formation, strain-induced structural transformation and local surface reconstruction, and adhesion. From a physics point of view, it is determined by short- and long-range interactions between the surfaces. However, the underpinning mechanism of friction and the upscaling from atomic phenomena to microscopic effects are still not understood. The classical friction laws were discovered by da Vinci and Guillaume Amontons, respectively, and were summarized much later by Charles Augustin Coulomb, who also contributed the third friction law. The three laws of friction describe that the friction force to resist sliding at an interface is (1) proportional to the normal force between the surfaces, (2) independent of the apparent contact area, and (3) independent of the sliding velocity. In this contribution it is shown that

nanocomposite diamond-like carbon (DLC) materials may show a breakdown of the Coulomb friction law.

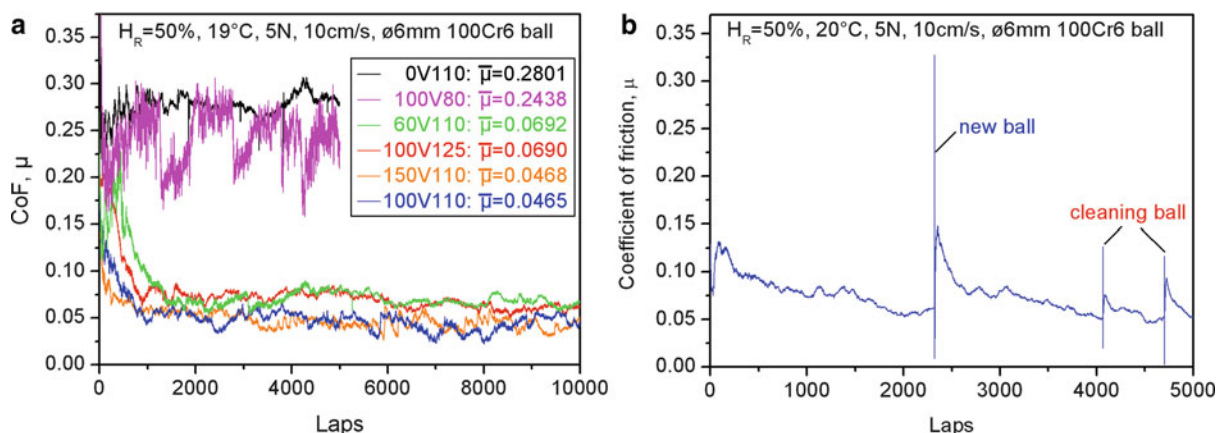
Friction of Nanocomposite DLC Coatings

A panoply of testing techniques has been developed to examine the tribological performance of PVD coatings. While the actual applications of DLC coatings in the automotive industry are often under lubricated contacts, tribological tests without lubricants and in ambient air or controlled atmospheres allow the investigation of the physical mechanisms that affect self-lubrication effects. The results of these tests do not always guarantee that a coating will be successful in a particular application, but they assist in determining the mode of wear of different coating materials in the stage of coating design and development. The well-defined loading conditions employed in laboratory tests may be different from those of real applications where lubricant and even abrasive particles are often present. Indeed, abrasive wear resistance is of paramount importance in many tribological applications, and so is the interaction of the oil and oil additives with the coating material. The temperature of the coated components in the contact is an important factor influencing the tribological behavior of the coatings. The thermal stability of the substrate material is also important in this sense in that the substrate needs to maintain its mechanical properties to properly support the coating in the contact.

DLC-based coatings may exhibit low friction, which has been attributed to the possible formation of a transfer layer on the surface of the uncoated counterpart. Recent research has paid some attention to the effects of transfer films on friction behavior, but their importance is still rather overlooked. The friction behavior of hydrogenated and hydrogen-free DLCs differs in the formation mechanism of transfer films within the contact. The former makes the contact between two similar hydrophobic DLC surfaces and in the latter a graphitic layer acts as solid lubricant. The influence of environment on the friction of DLC-based materials is a topic of controversy and remains under continuous debate. Contradictory reports can be found of the effects of adsorbed gases on the friction coefficient. For example, Zaïdi et al. observed that the steady-state coefficient of friction (CoF) of graphite fell as the partial pressure of oxygen gas increased or as the sliding velocity decreased (Zaïdi et al. 1990). In contrast, Heimberg et al. noted that adsorbed gases appeared to increase the CoF of hydrogenated DLC films (Heimberg et al. 2001). Obviously, the surface characteristics of materials play a crucial role in the tribological performance. The influence of the adsorption

of N_2 gas by DLC material during dry sliding experiments, causing a variation of friction coefficients with sliding speed and with time, has been explained by some authors through the use of the Elovich equation. The same equation can also describe our experimental results if only the adsorption of water vapor is considered. The friction behavior of TiC/DLC nanocomposite coatings is more complicated compared with pure DLCs. It has been shown that the presence of TiC nanocrystallites influences the formation of transfer layer, and in the case where the DLC matrix cannot efficiently shield TiC particles in the transfer films, as in the case of nanocomposite coatings with a large volumetric fraction of TiC nanocrystallites, TiC nanoparticles may hamper the formation of the transfer layer.

It is possible to deposit hydrogenated *nc*-TiC/*a*-C:H coatings by a closed-field unbalanced magnetron sputtering in an argon/acetylene atmosphere, using a coating system, configured with two Cr targets and two Ti targets, each pair being vertically opposed. The detailed set-up of the coating system has been documented elsewhere (Pei et al. 2005). The acetylene flow rate and substrate voltage bias were varied each in the range of 80–125 sccm (standard cubic centimeters per minute) and 0–150 V, respectively, to obtain different C/Ti ratios and nanostructures in the coatings. The coatings are named in such a way that the numbers preceding the character “V” indicate the substrate bias (in volts), and those following the acetylene flow rate (in sccm). Representative graphs of friction coefficient versus running laps are shown in Fig. 1a, where three different kinds of friction behavior of the nanocomposite coatings are recognized (Tay et al. 2000). The coating 0 V110 exhibits a nearly constant CoF and the coating 100 V80 shows rather large fluctuations in the friction coefficient curve. The mean CoF of the coatings 0 V110 and 100 V80 is above 0.2 (i.e., much higher than that of the rest four coatings). The other four coatings show not only a low steady-state CoF, but also a quick drop in the CoF from an initially high value of about 0.2 at the beginning of sliding until the transition point where the steady state is reached. This behavior is attributed to the gradual formation of a transfer film on the counterpart surface during the early stage of a tribotest, which makes the contact in between two basically similar hydrophobic DLC surfaces that contribute to self-lubrication. Against different counterparts (i.e., sapphire, alumina, and bearing steel balls), only slight differences in the friction coefficient are observed on the coatings with self-lubrication. This may imply that the interfacial sliding actually takes place between the transfer films on the ball and the surface of



Multiplex Coatings, Fig. 1 (a) Graph of coefficient of friction versus number of laps of the coatings under dry sliding against 100Cr6 steel ball and (b) self-lubricating effect of coating 100 V110

the coating, rather than sliding between the surfaces of the counterpart and the coating.

To prove that the self-lubrication is induced by the formation of transfer films, experiments were done in such a way that a fresh surface area of the ball was placed on the same wear track after the stationary state was reached (Fig. 1b). The friction coefficient immediately jumped to a value that characterizes the friction between the fresh steel surface and the coating. Again, the friction coefficient dropped down quickly as the transfer films gradually covered the ball surface. Further experiments have been carried out to demonstrate the effect of cleaning the contact area of the ball by rinsing with ethanol and drying with dry N_2 . Cleaning resulted in small peaks in the graph of friction coefficient followed by another quick drop in friction. However, the friction coefficient at the maximum of these peaks is lower than that when a fresh surface of the ball was brought into contact with the wear track, which may be possibly attributed to partially adhered films that could not be cleaned completely.

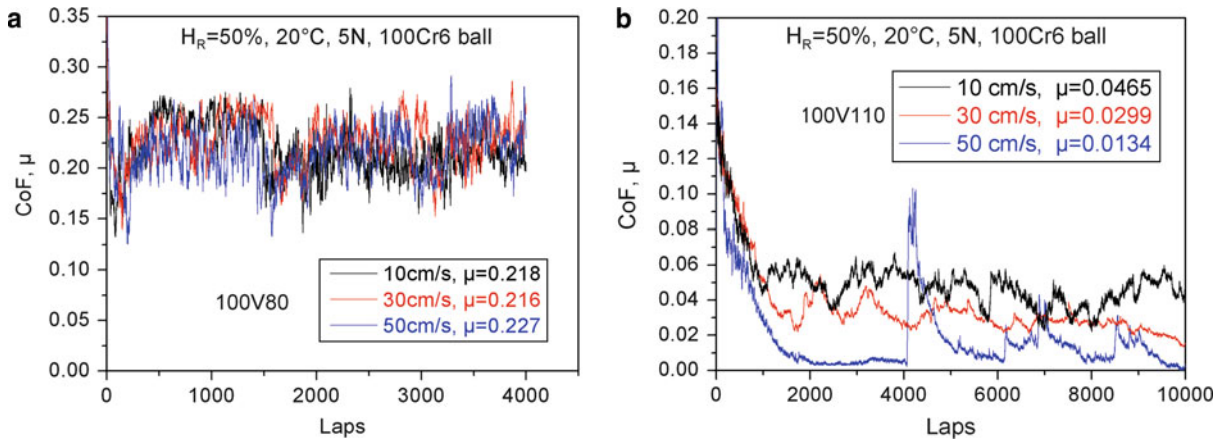
Influence of Sliding Velocity and Humidity

Figure 2 shows graphs of CoF versus laps for coatings 100 V80 and 100 V110, tested at different sliding velocities. The CoF graphs of coating 100 V80 in Fig. 2a are nearly horizontal curves with large fluctuations, and the mean values of CoF at different velocities are almost the same (i.e., 0.218, 0.216, and 0.227 at different sliding velocities of 10, 30, and 50 cm/s, respectively). Apparently, the coating 100 V80 without a self-lubricating effect exhibits a CoF that is independent of the sliding velocity. In other words, the Coulomb friction law holds in this case. In contrast, the CoF of coating 100 V110 drops

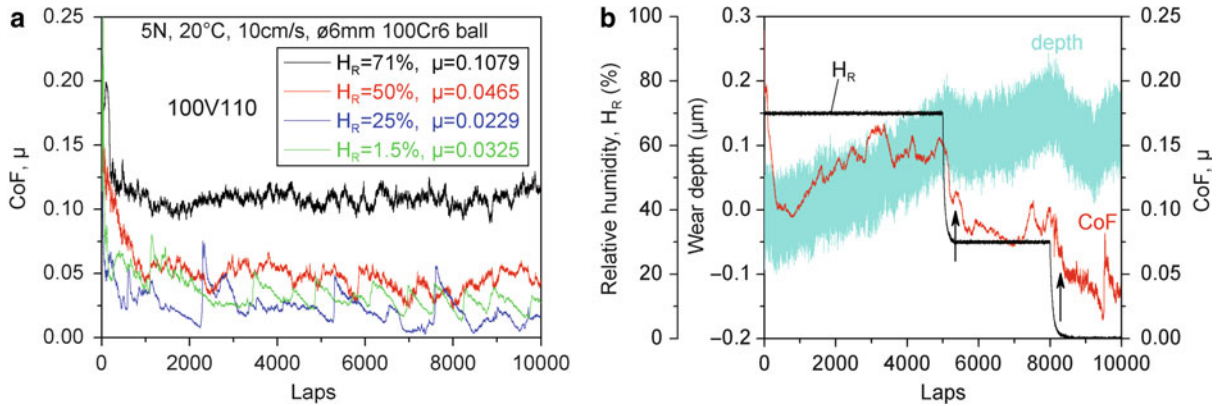
quickly from an initial high value of about 0.2 at the beginning of sliding to a very low value of CoF (<0.05) at the steady state, which is attributed to self-lubricating effects. A strong dependence of the steady state CoF on the sliding velocity is observed such that the faster the sliding velocity, the smaller the CoF (Fig. 2b). The steady state CoF of coating 100 V110 at sliding velocities of 10, 30, and 50 cm/s is 0.047, 0.030, and 0.013, respectively. It is clear that the Coulomb friction law is no longer valid when self-lubrication occurs, leading to jerky-type behavior.

The effects of relative humidity on friction are demonstrated in Fig. 3a. In general, the CoF of coating 100 V110 decreases with decreasing humidity. However, peaks have been recorded in the CoF curves when the relative humidity is equal to 25% or lower and they occur more often in dry air if the sliding velocity stays constant. This implies that the sliding velocity used is close to the limit in dry air and also raises the steady-state CoF slightly. These CoF peaks are attributed to the frequent breakdown of the transfer film as described above. Such a decrease of CoF with humidity has also been observed for coating 100 V125, where only a couple of peaks in the CoF curve occurred in dry air so that a monotonic decrease of CoF with decreasing humidity was recorded. By comparing the frequency of the CoF peaks in dry air between the coatings 100 V110 and 100 V125, it should be pointed out that the critical sliding velocity is also affected by the volumetric fraction of α -C:H matrix: the wider the TiC particle separation in the matrix, the higher the critical velocity will be.

The dynamic response of CoF to humidity and the corresponding change in the transfer film thickness are shown in Fig. 3b. Three levels of relative humidity



Multiplex Coatings, Fig. 2 Influence of sliding velocity on the coefficient of friction of the coatings: (a) 100 V80 and (b) 100 V110 under dry sliding against 100Cr6 steel ball



Multiplex Coatings, Fig. 3 (a) Effects of relative humidity on the CoF of the coating 100 V110 and (b) CoF dynamic response to humidity of coating 100 V125

(i.e., 70%, 30%, and 0%) were employed in a single-run tribotest run, where the transitions from high to low humidity were quickly realized by purging dry air into the testing chamber of the tribometer. It is interesting to note the change of CoF over the humidity transition periods. The CoF drops immediately once the humidity falls. Although the drop of CoF continues after the transition periods, there is a small step in the CoF decay corresponding to the end of the humidity drop, as marked by arrows in Fig. 3b, after which the slope of the CoF drop is further reduced. This points to different mechanisms involved in the reduction of friction. Thickening of the transfer film starts immediately after the humidity drops and lasts much longer than the transition periods, indicated by the segments of depth curve with a negative slope

as shown in Fig. 3b. Apparently, this thickening contributes to the whole course of the CoF drop. On the other hand, the thickness and coverage of adsorbed water molecular layer on the fresh transfer film and wear track are determined by the relative humidity and decrease as humidity lowers during the transition periods. It is understood that energy dissipation to the water molecular layer will accordingly decrease in the transition periods until a lower level is reached at lower humidity. This transition period is reflected in the first steep decline of CoF before the steps.

Adsorption and Friction

A lower humidity and a higher sliding velocity have similar effects on the frictional behavior of the self-lubricating

coatings, that is, they result in the formation of thinner transfer films that provide a lower friction coefficient. It is expected that both dependencies originate from the same microscopical mechanism – the influence of the water vapor condensation on the rheology of the transfer layer. The surface coverage of the adsorbed gas on a solid surface as a function of time (t) can be successfully described empirically using the so-called Elovich equation (Elovich and Zhabrova 1939):

$$\frac{dq}{dt} = Ae^{-\alpha q} \quad (1)$$

which can be integrated as

$$q = \frac{1}{\alpha} \ln(\alpha A) + \frac{1}{\alpha} \ln\left(t + \frac{\beta}{\alpha A}\right) \quad (2)$$

where α is a constant associated with the number of available adsorption sites over the surface and A is a constant related to the flux of adsorbing gas, which was found to be proportional to its partial pressure in certain cases of non-dissociative adsorption. β is determined by integration of (2) from the initial condition (t_i, q_i) to the situation of (t, q), that is,

$$e^{\alpha q_i} - A\alpha t_i = e^{\alpha q} - A\alpha t = \beta \quad (3)$$

When at the onset q_i and t_i are taken to be equal to zero, β is equal to unity. In a ball-on-disk tribo-test, each time the ball counterpart passes a point on the circular wear track, it “wipes” a contact area. Thereafter, the contact area is re-exposed to gases in the environment for new adsorption. The exposure time (t) between two successive wipes is inversely proportional to the sliding velocity (v):

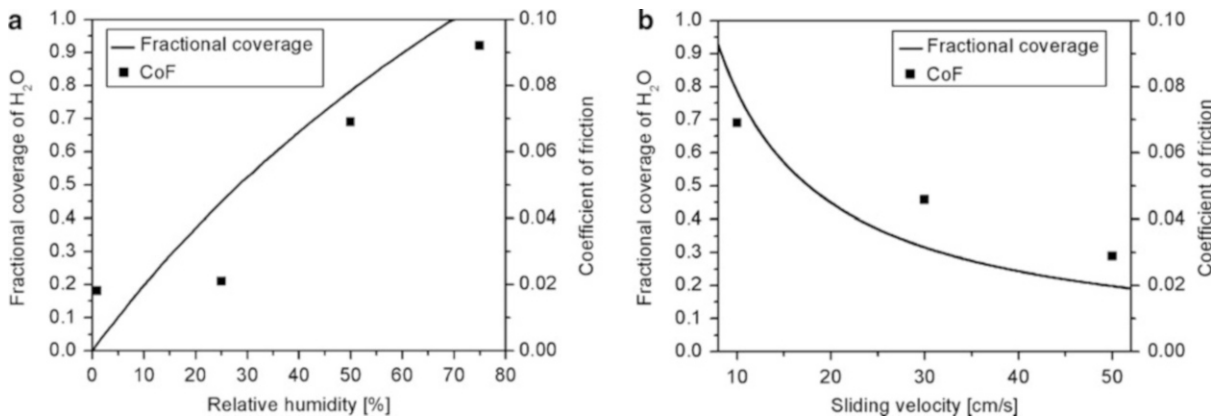
$$t = \frac{2\pi \cdot r}{v} \quad (4)$$

where r is the radius of the wear track.

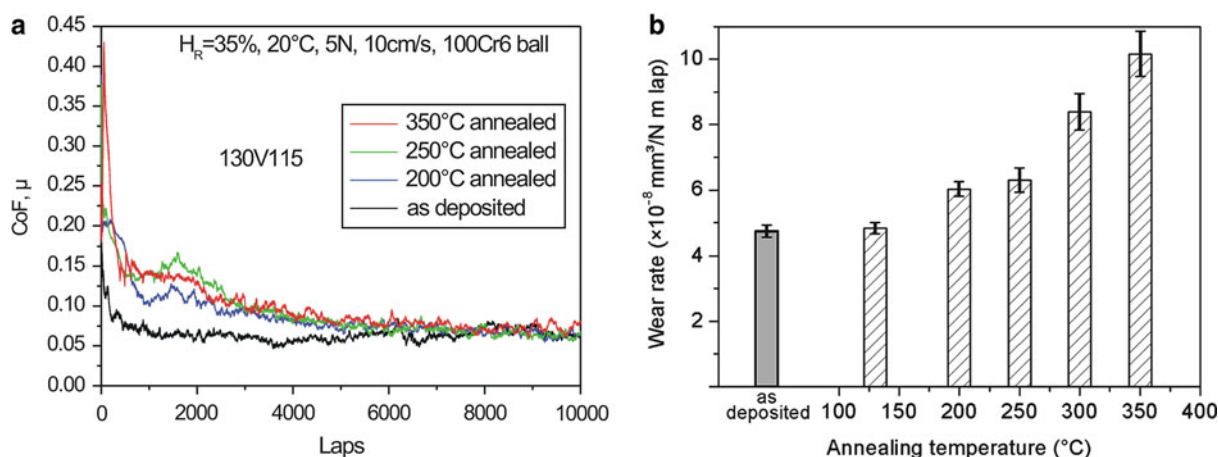
The Elovich equation was employed successfully to interpret the dependency of the friction coefficient of DLC coatings on the exposure time to a dry N_2 atmosphere (Heimberg et al. 2001). It is clear that the parameter governing the frictional behavior of the nanocomposite coatings is the relative humidity, which is directly proportional to the partial pressure of water vapor in the testing atmosphere. The fractional coverage of water vapor on the coating surface during the ball-on-disk test can be estimated as a function of the relative humidity and the sliding velocity, respectively (see Fig. 4). The experimental results indicate that the adsorption of water vapor is responsible for the frictional behavior, supported by the fact that the coefficient of friction of the coatings changes in a consistent trend with the fractional coverage of water vapor.

Influence of Temperature on Tribological Performance

In order to reveal the thermal resistance of the coatings, the variation of their hardness and elastic modulus was monitored by nanoindentation experiments after annealing them for 1 h at various temperatures in air. The DLC coatings are thermally stable up to 250°C but start to soften at 300°C. The softening of annealed coatings becomes pronounced above 350°C. In contrast, it is known that the mechanical properties of bearing steels (e.g., 100Cr6) degrade at a temperature of about 190°C. Figure 5 demonstrates the tribological performances of the



Multiplex Coatings, Fig. 4 Comparison between the CoF of coatings and the fractional coverage of water vapor on wear tracks calculated according to (2): (a) influence of relative humidity at sliding velocity of 10 cm/s and (b) effect of sliding velocity at 50% relative humidity



Multiplex Coatings, Fig. 5 Influence of annealing temperature on the tribological properties of coating 130 V115: (a) coefficient of friction (CoF) and (b) wear rate, tested at room temperature and 35% relative humidity, 5 N normal load and 10 cm/s sliding velocity against 100Cr6 ball

as-deposited and annealed coatings, respectively, as tested with ball-on-disk dry sliding against $\phi 6$ mm 100Cr6 balls under 5 N normal load and 10 cm/s sliding speed.

The as-deposited coatings show the typical self-lubricating behavior observed. That is to say, the CoF drops from an initially high value of about 0.2 at the beginning of sliding to a low steady-state CoF of about 0.06–0.08, after a transition period during which transfer films gradually form over the surface of the ball counterpart. The wear rate of the annealed coatings remains quite stable up to the annealing temperature of 250°C and increases significantly when annealed at temperatures above 300°C. In addition, the coatings annealed at higher temperatures exhibit a higher CoF at the onset of a tribotest and need longer to reach the low steady state CoF. For instance, the peak CoF at the beginning of sliding for the coatings annealed at 350°C is twice as high compared with the values measured for coatings annealed below 250°C. In other words, annealing at higher temperatures deteriorates the self-lubrication effect.

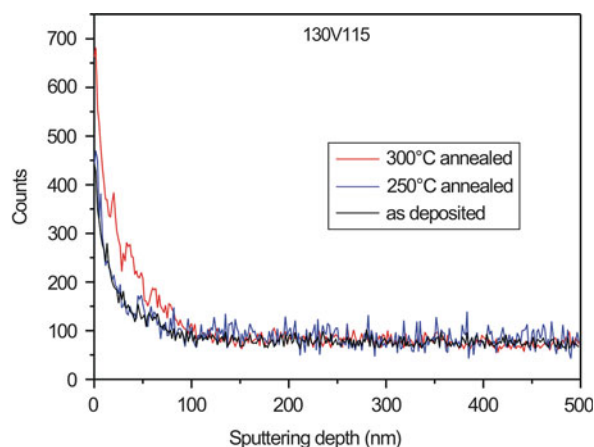
Auger depth profiles of the oxygen concentration in the as-deposited and annealed DLC coating is shown in Fig. 6 and reveals that oxidation of TiC/DLC coatings during 1 h annealing in air is negligibly small up to the annealing temperature of 250°C. A substantial increase in oxygen content has been detected after annealing at 300°C. However, the diffusion depth of oxygen is limited to about 100 nm in the annealed coating. Annealing for a longer time or at higher temperatures will increase further the oxygen content as well as the diffusion depth in the coatings, which may result in heavier degradation of

the annealed coatings. This oxidation behavior of a thin surface layer explains the higher initial CoF and the longer transition period needed to reach the steady state CoF for the coatings annealed at higher temperatures. Indeed, a CoF value as high as 0.28 was observed for the DLC coating that contains 16 at.% oxygen.

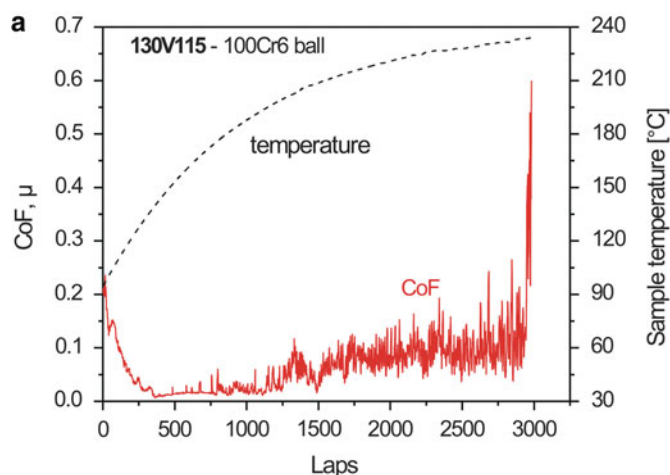
Ball-on-disk tribotests were performed starting at room temperature and gradually increasing the temperature during the test. The ultra-low friction regime of TiC/DLC coatings may maintain up to 200°C, as shown in Fig. 7, where the CoF (~ 0.01) is lower than the steady state CoF of the coating tested at room temperature. The friction coefficient gradually increases to a value of 0.1 as the temperature rises above 200°C. Thereafter, the coating suddenly fails at about 230°C with a sliding life shorter than 3,000 laps (revolutions), indicated by a peak in the friction trace up to a value of about 0.6. The latter is a typical CoF value for dry sliding between metal-to-metal contacts. This indicates that coating delamination is the failure mechanism, as confirmed by SEM observations in Fig. 7b.

The two different levels of CoF before abrupt failure of the coating deserve special comments. At the early stage of a test the CoF drops to a very low value of about 0.01 as the sample temperature rises rapidly towards 100°C. This phenomenon is attributed to the gradual desorption of water molecules from the sliding surfaces with increasing temperature, whose effect is equivalent to that of decreasing the humidity level of the atmosphere. When the sample temperature is above 200°C, the coating surface oxidizes locally at the hot spots on sliding contact where

the peak temperature may be well above 350°C. This leads to an increase in oxygen content in the top atomic layers of the coatings and consequently a significant increase in CoF, close to the CoF value of the annealed coatings. In addition, the 100Cr6 ball counterpart starts to soften at such high temperatures, which may destroy the transfer films and lead to adhesive wear due to the higher friction coefficient. Ultimately it results in delamination of the coating, as observed in Fig. 8b.



Multiplex Coatings, Fig. 6 Auger depth-profiling of oxygen element in the coating 130 V115 after annealing at different temperatures in comparison with that of the as-deposited coating

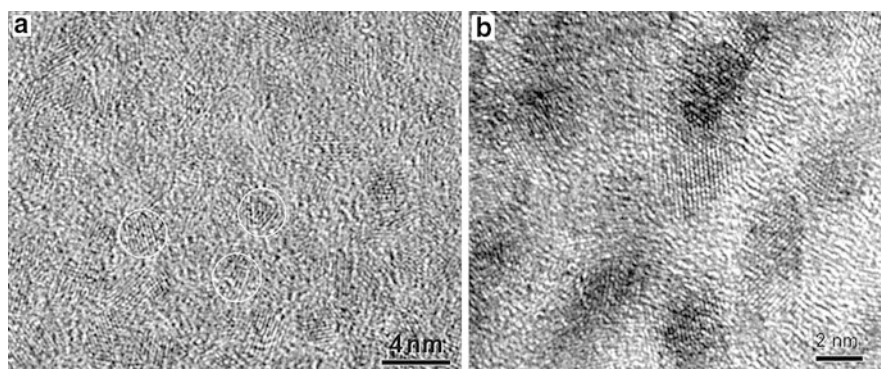


Multiplex Coatings, Fig. 7 Elevated temperature tribo-testing results of the coatings: (a) 130 V115 tested in air of 50 % relative humidity, 5 N normal load, 10 cm/s sliding velocity. (b) SEM micrograph of the middle part of the wear track on the coating 130 V115 showing a damaged spot where the coating delaminated during the elevated-temperature tribo-test. An arrow indicates the moving direction of the ball counterpart

It is known from the literature that DLC-based coatings are stable under annealing in air up to temperatures of at least 250 °C (Tay et al. 2000). Nevertheless, it has been shown that a degradation mechanism is present during the elevated temperature tribotests that limit their use to lower temperatures. The introduction of alloying elements in H-free DLC coatings reduces their hardness and their wear resistance, but decreases considerably the sliding friction coefficient at ambient temperature. Chemical alloying is therefore an efficient approach to vary the properties of DLC coatings for different applications. Nevertheless, care should be taken when selecting DLC coatings for applications at elevated temperatures. Indeed, in the higher temperature range all coatings provide lower friction coefficients than typically observed at ambient temperature. Pure DLC coatings are able to provide a sliding friction coefficient below 0.1 during sliding at high temperature up to 285°C, while the presence of Ti and Si alloying elements may decrease the maximum operating temperature of the coatings to 160°C.

Influence of Nanostructure and New Developments

In nanocomposite coatings composed of hard nanograins and a compliant matrix, two different designs have been recently put into practice in thin film applications, namely superhard and supertough nanocomposite coatings (Cavaleiro and De Hosson 2006). The concept of superhard nanocomposite coatings is based on the suppression of dislocation operation by using 3–5 nm small



Multiplex Coatings, Fig. 8 HR-XTRM micrographs showing (a) homogeneously distributed TiC nanocrystallites in TiC/a-C:H nanocomposite coating 100 V110 and (b) aligned TiC nanocrystallites separated by amorphous carbon matrix sub-layers of tunable thickness in a TiC/a-C nanocomposite coating (the multilayers are tilted about 60° to the bottom border of the micrograph)

grains and inducing grain incoherence strains with <1 nm thin matrix for grain separation. On the other hand, nanocomposite coatings generate a high density of inter-phase interfaces that assist in crack deflection and termination of crack propagation. The introduction of amorphous matrix may facilitate grain sliding that releases the strain energy stored. Based on these toughening mechanisms, the concept of supertough nanocomposite coatings has been proposed. For super toughness, a coating structure consisting of 10–20 nm nanocrystalline grains separated by 2–10 nm amorphous matrix has been suggested (Voevodin and Zabinski 2003). However, such a combination of large nanograins with a wide matrix separation is not favored regarding atomic migration during deposition. Magnetron sputtering with a negative substrate bias not higher than 150 V and without intentional substrate heating is less likely to generate atomic displacement over a distance of several nanometers through surface diffusion or subsurface migration. Therefore, many nanograins with sizes smaller than 10–20 nm are formed in the amorphous matrix with separations that are determined by the elemental concentration. In recent work on TiC/a-C:H nanocomposite coatings (Tay et al. 2000), the minimum size of TiC nanocrystallites observed is 2 nm. The particle size does not increase until the matrix separation diminishes with increasing Ti content to a certain limit, due to the competition between grain nucleation and grain growth. Based on these observations, a toughening mechanism of delocalizing cracks has been proposed. That is to say, keeping the suppression of crack nucleation in an amorphous matrix, the introduction of nanoparticles may spread the localization of cracks into a delocalized state. The ductility and therefore the

toughness will be enhanced provided the particle size becomes of approximately the same size as the separation.

As is clear from the aforementioned results the size and separation of TiC nanocrystallites in the DLC-based nanocomposite coatings are crucial for tribo-functionality. Self-lubricating effects have been only observed on the TiC/a-C:H nanocomposite coatings once the separation of the nanocrystallites in the a-C:H matrix is sufficiently wide, that is to say, a mean separation distance close to the size of the TiC nanoparticles. In fact nanocrystalline TiC (*nc*-TiC) particles may serve as a promoter for surface graphitization of the a-C:H matrix that leads to ultra-low friction. Surface graphitization of the a-C:H matrix is boosted due to the high localized shear stresses applied by the exposed *nc*-TiC in the transfer films (nano-scaled asperities). On the other hand, these TiC nanocrystallites also scratch the coating surface and facilitate wear. As a result, there is a trade-off between CoF and wear rate towards the low volume fraction of *nc*-TiC. It is thus understandable that the nanocomposite coatings exhibit even smaller CoFs than those of pure DLC coatings (typically 0.1 ~ 0.15 under comparable conditions of loading and counterpart), where such hard and sharp nanoscaled TiC asperities are missing. The interesting point here is that various combinations of CoF and W_R can be selected for different applications according to whether a high wear resistance or a low friction is the principal objective.

A new development is highlighted in Fig. 8, showing that a nanocomposite coating can be designed of homogeneously distributed TiC nanocrystallites or aligned TiC nanoparticles. To manipulate size and separation of TiC nanocrystallites, a possible approach is to tailor a

multilayered structure into a nanocomposite coating, such that the nanocrystallite containing sub-layers are separated by the matrix sub-layers of desired thickness via composition fluctuation during deposition. Through such an integrated multilayer-composite structure, the size and separation of TiC nanoparticles can be controlled independently (Fig. 5b), which was not possible in a homogeneous nanocomposite coating. It turned out that the multilayered nanocomposite coating exhibits essential enhancement in fracture toughness and it is hardly possible to induce radial cracks via nanoindentation with a cube-corner diamond indenter even if the indenter has penetrated through the entire coating thickness. In particular, its tribological performance is also superior and comparable with the advanced TiC/a-C:H nanocomposite coatings (Pei et al. 2006). Further detailed characterization of the nanostructure and properties of TiC/a-C nanocomposite coating are published elsewhere (Pei et al. 2008; Chen et al. 2009). TiC/hydrogenated DLC coatings are stable after annealing in air up to a temperature of at least 250°C. No variation of their mechanical properties or appreciable oxidation was observed. Its wear resistance also remained unchanged up to this temperature. Nevertheless, tribological tests at high temperature resulted in failure of the self-lubrication mechanism and subsequent coating failure at temperatures around 200°C.

Key Applications

From an industrial point of view, during the 1980s the applications of the Me-DLC coatings were very limited. The company Balzers tried to commercialize the Me-DLC coatings and in particular the W-C:H, but no large production volumes were achieved. It was not until approximately 1994–1995 that the first large-scale industrial production of the Me-DLC coating was started by the company of Robert Bosch GmbH. The application concerned parts for the first generation of high-pressure diesel fuel injection systems for engines and the coating used was a W-C:H coating. The coating is used on critical components in the injection system to avoid scuffing and abrasive/adhesive wear due to the very high injection pressures used in, for example, common rail diesel fuel injection systems. PVD and PA-CVD coatings are widely used today in the diesel fuel injection systems and have become a necessity for survival of critical components when the injection pressure is rapidly approaching 2,000 bar. No diesel fuel injection system would today survive without PVD and PA-CVD coatings.

It is expected that the demand for tribological coatings will grow for industrial applications in the near future.

There are several factors contributing to this development. One important factor is the rapid development of engine technology in the automotive industry over the last years. Today, the demand for components in an engine leads to higher precision in the tolerances for machining. The growing precision for components is not only demanded when the engine is brand new, but also over the lifetime of the engine. One of the main reasons for this is the need to control all parameters in the combustion accurately to lower fuel consumption and emissions from the engines. In general, components today are also expected to have a longer lifetime with less maintenance intervals. This has significant impact on the wear situation of many components in the engine. Another important driving force for coating engine components with PVD coatings is the increasing power density components are exposed to in modern engines, and here especially the diesel engine plays an important role. The introduction of the common rail diesel fuel injection system in 1997 has led to an ever-increasing need to protect components by tribological coatings, not only in the diesel fuel injection system itself, but also in many other areas in the engine. The reason for this development is the increased injection pressures of diesel that results not only in cleaner combustion and less fuel consumption but also in higher power and torque of the engine.

The injection pressure of today's generation of common rail is approximately 1,800–2,000 bars and the next generation that will be introduced in a couple of years' time will exceed 2,000 bars. At these extremely high pressures, many components in the diesel fuel injection system cannot survive without protective DLC-type coatings to prevent, for example, abrasive wear and seizure of the components. The increased power and torque also creates a significantly higher power density in many power train components, since it is not permissible to design bigger or heavier components in modern engines due to space and weight saving. On the contrary, many components are today both more highly loaded and at the same time smaller and lighter. Critical areas include valve train components like tappets and rocker arms, piston rings and pins in the piston assembly, and journal bearings in the crankshaft area.

To tackle these problems, automotive engine design engineers try to use primarily conventional surface engineering technologies like higher quality steels, better performing heat treatments, and improved surface finish by the use of better grinding and polishing techniques. The reason for this is mainly cost and caution regarding introduction of new technology in mass production. Only when conventional technologies no longer work for the function of a component is alternative surface engineering

considered. However, there are a few exceptions to the above discussion. In some industrial applications, there has been an opportunity to replace the steel material of the original part with cheaper steel with a PVD coating deposited on top. The end result has been a component that performs better and is cheaper than the original part. This is, of course, the ideal situation for applying a PVD coating.

The results presented here show that it is possible to tailor coating properties to accommodate the need for different industrial applications. For example, in an industrial application like roller bearings it is important that the coating is low stressed and has an excellent fatigue resistance to being able to survive in the application and provide the expected extension in lifetime compared with uncoated roller bearings. The results indicate that a relatively soft and multilayered Me-DLC coating would be suitable for this type of application. The combination of low coefficient of friction with moderate hardness and stress in combination with an excellent fatigue resistance indicates that it could perform well. Another example is the industrial application of coatings on diesel fuel injection parts. The plunger used in the pump for building up the very high injection pressure (up to 2,000 bars in the latest generation of common rail systems) is typically coated. The requirements for this application are to reduce scuffing and seizure of the plunger to the very tight bore. Much abrasive wear is normally also present, since the fuel might be contaminated with particles. The typical coating used to solve these problems in the application is nanocomposite DLC based.

Chemical effects, in particular, should be considered if the influence of the atmosphere on the coefficient of friction must be explained. For a-C:H a low humidity atmosphere is the preferred condition for low friction. The H-terminated surfaces of both counterparts ensure their contact occurs under low adhesion, so that the transfer layer will be kept at the optimal thickness. The presence of humidity influences the surface properties of the counterparts, increasing their adhesion. Under these conditions the thickness of the transfer layer will vary under sliding contact, with subsequent increase of friction, which will increase wear, modifying the transfer layer thickness and leading to an unstable situation that will finally lead to high-friction sliding. For H-free a-C the situation is different, in that in this case the transfer layer formation is “dynamic,” with graphite plates continuously transferring between coating and ball, because of the low shear rate along the basal planes of graphite. The ease of mutual sliding of the graphite basal planes can be improved with the presence of intercalated water

molecules, giving a very different behavior as compared with a-C:H. Also in this case, when the transfer layer thickness has reached an optimum thickness there will be a stable situation, as the corresponding low friction will ensure no further modifications of its thickness will occur. In this case sliding in vacuum or inert atmosphere leads to an unstable high-friction state. Under vacuum sliding, the p_z orbitals of each graphite atom will be dangling on the surface unsaturated, which will increase the adhesion between two such surfaces enormously, leading to high friction and wear and no possibility for transfer layer formation. Chemical effects are probably the reason why lamellar, low-shear strength materials such as Ti_3SiC_2 do not exhibit low friction following the formation of a transfer layer (El-Raghy et al. 2000), which is in contrast with the classical theory of low friction, which states that a low shear-strength material on top of a hard substrate is the desired low-friction configuration. To further support the framework presented here, the behavior of polymers such as HDPE (high-density polyethylene) and PTFE (polytetrafluoroethylene) sliding against glass can be mentioned (Hutchings 1992). These polymers form a transfer layer on the hard counterpart, but their initial coefficients of friction remain around 0.2–0.3, while the transfer layers are micrometers thick. As the sliding progresses the transfer layers become much thinner, and only then coefficients of friction as low as 0.05 are measured (for PTFE).

The following mechanism may be proposed to explain the jerky-type frictional behavior of the nanocomposite DLC coatings. The transfer film and the wear debris accumulated in front of the wear scar of the ball counterpart actually carry the contact load; their rheology is of prime importance in lowering the friction. When this nanosized debris is (in dry air) not covered by water molecules, they may be brought into (and easily sheared at) the sliding interfaces with a very weak interaction between themselves as well as between the surface of the wear track. However, at high sliding velocity the flow of debris into the sliding contacts may be interrupted due to the collapse of accumulated debris in front of the wear scar. Such collapses lead to the jerky-type behavior due to the frequent breakdown of transfer films and correspondingly to the peaks of CoF beyond the critical sliding velocity. At faster sliding velocity and/or lower level of humidity the films become thinner and looser and therefore they easily break. Collapses of the accumulated debris are expected to be more damaging to the thinner transfer films seen at higher sliding velocities. Condensation of water molecules collected from the surface of the wear track may change the nature of transfer films in humid air. It is well known that

adsorbed gases, especially water vapor, increase the rate of densification of particulate materials, for instance in the wear debris accumulated here. Therefore denser and thicker transfer films are formed at higher humidity and at lower velocity. One can conclude that it requires more energy mechanical work to slide and smear such a film than a loosely compacted and grainy one between the sliding surfaces.

Acknowledgments

The authors acknowledge financial support from the Materials Innovation Institute—the Netherlands and the Foundation for Fundamental Research on Matter (FOM-Utrecht).

References

- A. Cavaleiro, J.T.M. De Hosson, *Nanostructured Coatings* (Springer, New York, 2006)
- C.Q. Chen, Y.T. Pei, K.P. Shaha, J.T.M. De Hosson, *J. Appl. Phys.* **105**, 114314 (2009)
- J.T.M. De Hosson, O. Kanert, A.W. Sleswijk, in *Dislocations in Solids*, ed. by F.R.N. Nabarro (North-Holland, Amsterdam, 1983), pp. 441–534
- S.Y. Elovich, G.M. Zhabrova, *Zh. Fiz. Khim* **13**, 1761 (1939)
- T. El-Raghy, P. Blau, M.W. Barsoum, *Wear* **238**, 125 (2000)
- J.A. Heimberg, K.J. Wahl, I.L. Singer, A. Erdemir, *Appl. Phys. Lett.* **78**, 2449 (2001)
- I.M. Hutchings, *Tribology: Friction and Wear of Engineering Materials* (Edward Arnold, London, 1992). Co-published by CRC Press, Boca Raton, FL, USA
- J.W.J. Kerssemakers, J.T.M. De Hosson, *J. Appl. Phys.* **83**, 3444–3445 (1998)
- Y.T. Pei, D. Galvan, J.T.M. De Hosson, *Acta Mater.* **53**, 4505 (2005)
- Y.T. Pei, P. Huijzen, D. Galvan, J.T.M. De Hosson, *J. Appl. Phys.* **100**, 114309 (2006)
- Y.T. Pei, C.Q. Chen, K. Shaha, J.T.M. De Hosson, J. Bradley, S. Voronin, M. Cada, *Acta Aeta Materialia* **56**, 696–709 (2008)
- B.K. Tay, D. Sheeja, S.P. Lau, X. Shi, B.C. Seet, Y.C. Yeo, *Surf. Coat. Technol.* **130**, 248 (2000)
- A.A. Voevodin, J.S. Zabinski, *Thin Solid Films* **370**, 223 (2003)
- H. Zaïdi, D. Paulmier, J. Lepage, *Appl. Surf. Sci.* **44**, 221 (1990)

Multi-Row Bearings

XIAOLAN AI

Timken Technology Center, The Timken Company,
Canton, OH, USA

Synonyms

Double-row bearings; Four-row bearings; Heavy-duty bearings; Multi-row self-aligning bearings; Package bearings

Definition

Multi-row bearings are bearing assemblies with two or more rows of rolling elements arranged between raceways. Typical examples of multi-row bearings are spherical ball bearings, spherical roller bearings, and package bearings.

Scientific Fundamentals

For the purpose of increasing load capacity, bearings are constructed with two or more rows of rolling elements. Two-row configurations are also constructed for self-retaining, for preloading, for bi-directional axial load carrying, or for reduction of internal friction losses. The following sections outline commercially available multi-row bearings.

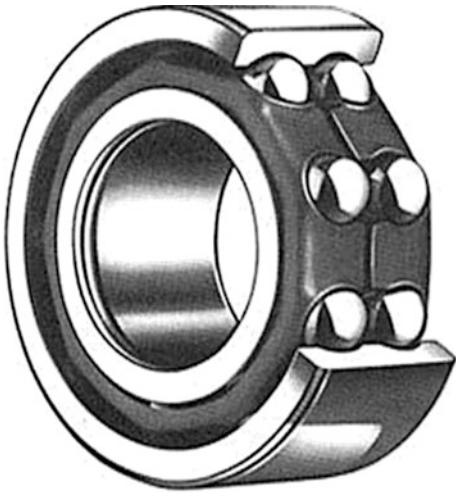
Double-Row Ball Bearing

Figure 1 shows a double-row deep-groove ball bearing. Both the inner and outer rings have two grooved raceways, one for each row of balls. The center distance of the two grooves on the inner ring is essentially the same as the center distance of the grooves on the outer ring. The bearing has two separate retainers for retaining two rows of balls, respectively. This bearing has greater radial load-carrying capacity than the single-row types. Load sharing is determined by the geometrical accuracy of the grooves. Double-row ball bearings perform similarly to single-row ball bearings.

The spacing between the centers of the two grooves on one ring can be controlled with respect to another ring to reduce the bearing axial clearance or to preload the bearing. This type of bearing is often classified as the double-row angular-contact ball bearing. When spacing between groove centers on the outer ring is greater than that of the inner ring, the bearing is referred to as the rigid type; when spacing between groove centers on the outer ring is smaller than that of the inner ring, the bearing is referred to as the non-rigid type. Double-row angular-contact ball bearings can carry thrust load in either direction or combined radial and thrust loads. Bearings of the rigid type are capable of handling moment load effectively. The bearings perform is essentially the same as duplex pairs of single-row angular-contact ball bearings.

Self-Aligning Double-Row Ball Bearing

Figure 2 shows a double-row self-aligning ball bearing. It contains an outer race ring having an outer raceway, an inner race ring having two grooved inner raceways, two rows of balls, and two separate cages each retaining a row of balls. The outer raceway of this bearing is a portion of a sphere. Thus balls on the inner raceways are internally aligned to accommodate any misalignment, free from any



Multi-Row Bearings, Fig. 1 A double-row deep-groove radial ball bearing



Multi-Row Bearings, Fig. 3 A double-row spherical roller bearing (Courtesy of the Timken Company)



Multi-Row Bearings, Fig. 2 A double-row self-aligning ball bearing

moment loads. Unlike conventional ball bearings, the outer raceway of self-aligning ball bearings does not conform well to the balls. This results in a reduced Hertzian contact area and load-carrying capacity (see Ball Bearing et al. 1981). For this reason, self-aligning ball bearings are constructed in a two-row configuration. Self-aligning ball bearings are used to carry primarily radial loads and a limited amount of axial loads in both directions.

Double-Row Spherical Roller Bearings

Most spherical roller bearings are in two-row configurations. Figure 3 shows a double-row spherical roller

bearing. It comprises an outer race ring, an inner ring having two raceways, and two rows of barrel-shaped rollers that are separated and retained by an integral cage. The outer race ring has a single outer raceway that is a portion of a sphere. This allows the inner ring and roller assembly to dynamically align inside the outer raceway, permitting a limited amount of angular displacement between bearing shaft and housing bore. Standard spherical roller bearings are designed to accommodate misalignment up to $\pm 1.5^\circ$. Rollers of a spherical roller bearing lie at an angle relative to the axis of the bearing. A seating rib is presented between the two rows of rollers. The seating rib can be a part of or floating around the inner one. The rollers are in either symmetrical or asymmetrical barrel shape, and conform closely to both the inner and outer raceways. The high degree of conformity makes spherical roller bearings suitable for heavy-duty applications. Because of the non-zero contact angle, spherical roller bearings are capable of carrying a certain amount of thrust load in either direction along with the radial load.

Unlike cylindrical roller bearings or tapered roller bearings, true rolling motion cannot be achieved at contacts between the rollers and raceways for spherical bearings (see Harris 2001). Thus, spherical bearings inherently have higher frictional torque than cylindrical bearings and are not suitable for high-speed applications.

Multi-row Cylindrical Roller Bearings

To achieve greater load carrying capacity, cylindrical roller bearings are often constructed in configuration of two or more rows. This is also done for the purpose of reducing

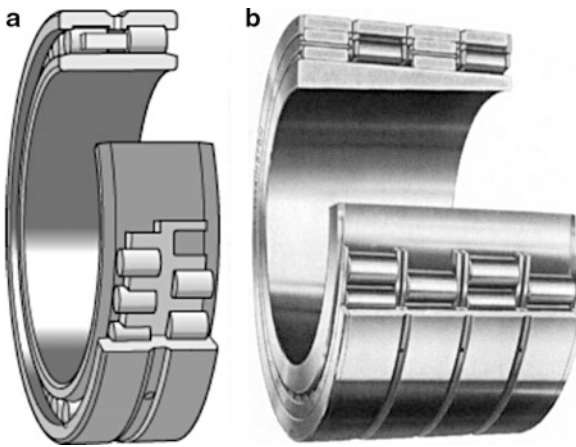
the tendency of roller skewing. Figure 4a illustrates a double-row cylindrical roller bearing for machine tool spindle application and Fig. 4b shows a multi-row cylindrical roller bearing for steel rolling mill application. As the number of rows of rollers increases, alignment becomes increasingly important and so is the lubrication. Ports and grooves are often added to the outer rings to facilitate lubricant circulation.

Multi-row Tapered Roller Bearings

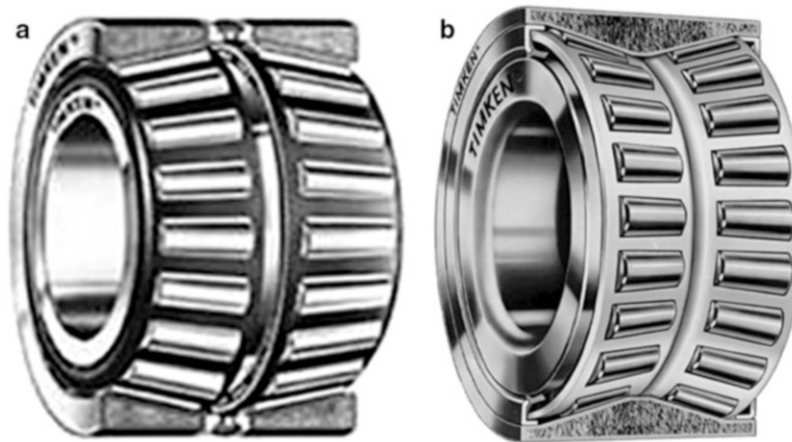
Tapered roller bearings are available in double-row and four-row assemblies. They are primarily used in applications where greater radial load carrying capacity is needed. The double-row tapered roller bearings are the most

common configuration in the multi-row tapered roller bearing family. It comes with either a single-piece double inner race ring or a single-piece double outer race ring arrangement. The single-piece double inner race bearing assembly has a single inner ring with double tapered raceways and two separate outer rings. It is usually supplied complete with an outer ring spacer as a preset assembly. This bearing type is often referenced to as TDI (Tapered roller bearing with Double Inner raceways), as shown in Fig. 5a. Applications of TDI bearings are found in gear reduction units, cranes, marine drives, and many other types of industrial equipment. The single-piece double outer race bearing has a single outer ring with double tapered raceways and two separate inner rings. It is usually supplied complete with an inner ring spacer as a pre-set assembly. This type of bearings is referred to as TDO bearings (Tapered roller bearing with Double Outer raceways), as shown in Fig. 5b. Because of the wide effective bearing spread, TDO bearings are capable of supporting overturning moments. TDO bearings find wide applications in gear reduction units, machine tools, cranes, and mining and construction equipment. Bearing internal clearance is controlled by precision ground spacers. A properly sized spacer can be selected or customized to meet the required bearing setting for a given application.

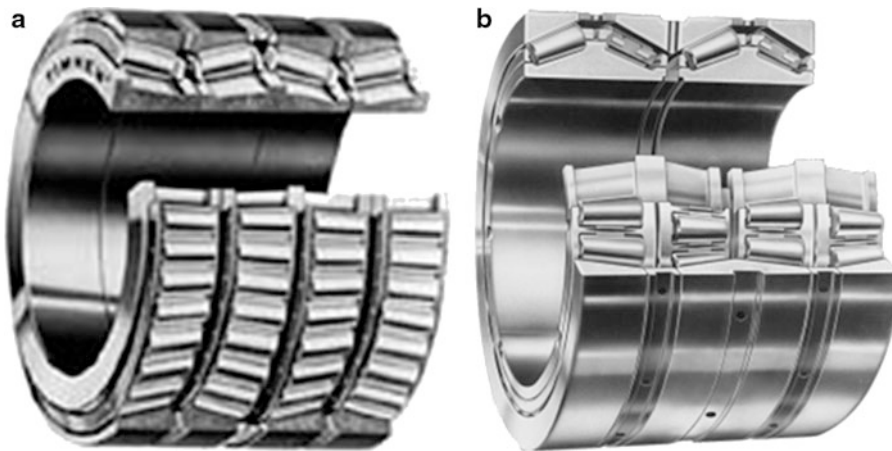
Four-row bearings are constructed either in two inner rings and three outer rings configuration, as illustrated in Fig. 6a, or in two outer rings and three inner rings configuration, as illustrated in Fig. 6b. In essence, four-row bearings combine the inherent high-load, radial/thrust capacity and direct/indirect mounting variables of



Multi-Row Bearings, Fig. 4 (a) A double-row cylindrical roller bearing (b) A four-row cylindrical roller bearing



Multi-Row Bearings, Fig. 5 (a) A double-row tapered roller bearing in TDI arrangement (Courtesy of the Timken Company). (b) A double-row tapered roller bearing in TDO arrangement (Courtesy of the Timken Company)



Multi-Row Bearings, Fig. 6 (a) A four-row tapered roller bearing with two inner rings and three outer rings (Courtesy of the Timken Company). (b) A four-row tapered roller bearing with two outer rings and three inner rings (Courtesy of the Timken Company)

tapered roller bearings into assemblies of maximum load rating in a minimum space. A typical application is found on the roll necks of rolling mill equipment. Almost all four-row bearings are provided with spacers as preset bearing assemblies. The spacers are ground and customer fitted to provide a bearing setting to meet application requirements.

Key Applications

Multi-row bearings are often used in applications where heavy load carrying capacity or higher support stiffness is required. Typical applications include steel rolling mills, machine tool spindles and vehicle wheel ends

Cross-References

- [Function and Structure of Rolling Element Bearings](#)
- [Radial Bearings](#)
- [Self-aligning Bearings](#)

References

- B.J. Hamrock, *Ball Bearing Lubrication, The Elastohydrodynamics of Elliptical Contacts* (Wiley, New York, 1981)
- T.A. Harris, *Rolling Bearing Analysis*, 4th edn. (Wiley, New York, 2001)

Multi-Row Self-Aligning Bearings

- [Multi-Row Bearings](#)

Multiscale Contact

- [Stochastic Contact Theories: Other Theories Based on the Greenwood-Williamson Model](#)

